Informatics Institute of Technology

In Collaboration With

The University of Westminster, UK

The University of Westminster, Coat of Arms

A Novel Approach to Time Series Forecasting using Liquid Time-constant Networks

A Literature Review by

Mr. Ammar Raneez

W1761196 | 2019163

Supervised by

Mr. Torin Wirasingha

October 2022

This Project Proposal is submitted in partial fulfilment of the requirements for the BSc (Hons) Computer Science degree at the University of Westminster.

Contents

List of Tables

List of Figures

Acronyms

AI	Artificial Intelligence.
API	Application Programming Interface.
ARIMA	Autoregressive Integrated Moving Average.
BPTT	Back-Propagation Through Time.
BTC	Bitcoin.
CT-GRU	Continuous-time Gated Recurrent Unit.
CT-RNN	Continuous-time Recurrent Neural Network.
DL	Deep Learning.
GPU	Graphics Processing Unit.
LSTM	Long Short-Term Memory.
LTC	Liquid Time-constant.
ML	Machine Learning.
(s)MAPE	Symmetric Mean Absolute Product Error.
MASE	Mean Absolute Scaled Error.
MSE	Mean Squared Error.
N-BEATS	Neural Basis Expansion Analysis for interpretable Time Series.
NER	Named Entity Recognition.
NLP	Natural Language Processing.
ODE	Ordinary Differential Equations.
POC	Proof-Of-Concept.
REST	Representational State Transfer.
RMSE	Root Mean Squared Error.
RNN	Recurrent Neural Network.
SOTA	State-Of-The-Art.
TS	Time Series.
SDE	Stochastic Differential Equations.

1. CHAPTER OVERVIEW

In this chapter, the author will attempt to critique related work within the domain of TS and BTC forecasting and work in the broader realm of open market forecasting, alongside providing a background on the chosen field. Moreover, the author has brought forward further research possibilities that could potentially open up.

2. CONCEPT MAP

Upon researching available literature, the scope of this literature review was broken down by a concept map. The map was designed to identify the required literature to be covered; it can be found in **APPENDIX A**.

3. PROBLEM DOMAIN

3.1 Time series forecasting

TS forecasting has attracted multiple researchers to develop a robust, scalable, and usable solution. It can play a crucial role in various real-life problems ranging from forecasting the weather to predicting traffic to predicting EEGs of patients in medical monitoring (Lara-Benítez, Carranza-García and Riquelme, 2021). In a nutshell, any problem with a temporal component can be considered a potential domain for applying TS forecasting, which is one of many reasons for so much demand and significant impact on business.

Until recent advancements in ML, traditional statistical models have been used with the help of domain expertise. However, with increasingly available data and computing power, ML has become vital in creating forecasting models for the next generation (Lim and Zohren, 2020). ML provides a means of learning temporal dynamics in a data-driven manner compared to their traditional counterparts (Ahmed et al., 2010).

In particular, DL has gained tremendous popularity in recent times for its remarkable accomplishments in image classification, NLP, and reinforcement learning, as it can learn representations from complex data without needing explicit engineering (Lim and Zohren, 2020).

3.2 Cryptocurrencies (Bitcoin)

Blockchain technology is based on peer-to-peer connectivity and cryptographic security that removes the need for a third-party centralized system, thereby demonstrating transparency and trust compared to traditional monetary systems (Rejeb, Rejeb and G. Keogh, 2021).

After the global financial crisis of 2008, BTC, a peer-to-peer electronic cryptocurrency, was introduced due to the loss of public trust in banking systems (Nakamoto, 2008). The creation of cryptocurrencies was motivated by the need to create a system that allowed simple, fast and cheap transactions that were not influenced by a third party (ex: banks) (Kfir, 2020). Moreover, multiple scholars and enthusiasts considered BTC the future of currency (Bouri et al., 2018).

Over time, over 1,600 cryptocurrencies (Wilson, 2019) have been introduced as a means of exchange of goods and services; this research focuses on BTC.

3.2.1 Opportunities of cryptocurrencies

Although cryptocurrencies do not solve all financial problems, they address many issues, such as the lack of trust, instability, and inefficient transactions (Nakamoto, 2008). Buhalis et al. (2019) stated that transactions are much more methodical and straightforward as they can be used to prevent fraudulent exchanges and payments.

Of many, the most critical use of cryptocurrencies is for online payments. The usage of credit cards and other cashless payments has given rise to the emergence of cryptocurrencies, which are becoming the most popular form of payment on the internet (Wang et al., 2019). Pournader et al. (2020) state that companies can make instant money transfers by utilizing this, thereby reducing commission.

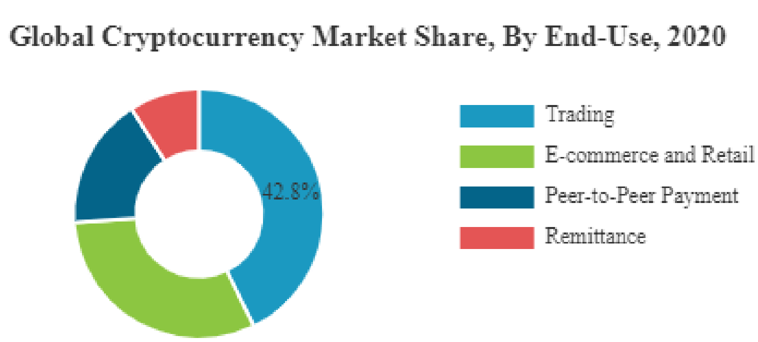

Figure 1: Global Market Share (Cryptocurrency Market Size, Growth & Trends | Forecast [2028], 2021)

Specifically, Nica et al. (2017) state that the popularity of BTC can be credited to the low transaction fees compared to traditional payment services. However, for this to be maintained, Alonso-Monsalve et al. (2020) mention that trading must occur on a "*basis of different assumptions that may not hold in specific situations, including quick links between users, low transaction costs, and high liquidity.*" Nakamoto (2008) further mentions that the settlement time of BTC is much faster than any non-cash transactions, which may take days or weeks.

The use of cryptocurrencies is increasing daily, and it may come to the point where it replaces traditional currencies as credibility grows. However, it comes with a fair number of challenges.

3.2.2 Challenges of cryptocurrencies

Given the absence of legislation and regulatory standards, certain risks have been introduced upon the rapid growth of cryptocurrencies that raise several concerns regarding virtual currencies' integration into the financial system (Avdeychik & Capozzi, 2018).

Although being decentralized has advantages, the absence of a governing authority has given rise to the development of online black markets. Due to its anonymity, it is even more challenging to trace the identity of a specific operation, user, or criminal activity (Baldimtsi et al., 2017). Kerr (2018) stated that BTC had been used as a tool for conducting business in the black market. Miller (2016) also mentioned that cryptocurrencies had been used to assist in selling drugs, weapons, other unlawful goods, and even child pornography. Therefore, the growth could threaten certain activities, people's incomes, and even their lives (Scharding, 2019).

Nevertheless, it is worth mentioning that upon the exchange of cryptocurrencies for fiat, source detection is much more accessible.

3.2.3 Why have cryptocurrencies taken the world by storm?

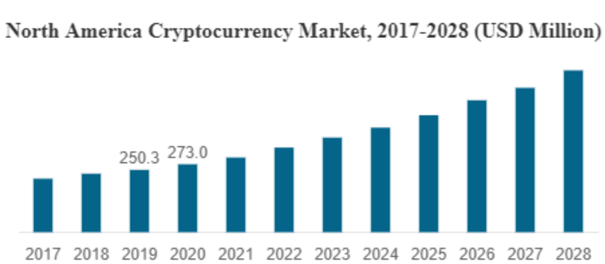

Figure 2: Latest Trends (Cryptocurrency Market Size, Growth & Trends | Forecast [2028], 2021)

As depicted in the above figure, the market of cryptocurrencies seems to increase each year steadily with the increasing popularity of digital currencies. The increasing popularity, in turn, gained the trust of the central bank, which established a patent, the Central Bank Digital Currency (CBDC), for digital currency projects across many developed countries. Multiple countries, including China, Thailand, Uruguay, and the Caribbean, have grown to support the CBDC for utilizing digital currency as a means of exchange, which could signify the explosion in popularity (Cryptocurrency Market Size, Growth & Trends | Forecast [2028], 2021).

The above growth could grow exponentially after more developed countries get involved.

3.2.4 Cryptocurrency exchanges

A cryptocurrency exchange is a platform that facilitates the trading, buying, and selling of cryptocurrencies. They also allow the conversion of cryptocurrencies to a fiat currency and further withdraw it to a local bank account (Want to Buy Crypto? Here's What to Look for In a Crypto Exchange, 2021).

Some of the more popular platforms include *Gemini*, *Coinbase* & *Binance*.

4. EXISTING WORK

4.1 Time series forecasting

Multiple methods have been introduced for the general domain of TS forecasting, ranging from traditional statistics to deep learning models. Since the domain of TS forecasting is more technologically specific, insights on the various approaches are broken down under the **technological review** section of this chapter.

4.2 Bitcoin forecasting

The domain of BTC forecasting has been gaining exponential traction within the past decade. Websites are making their historical data available to the public, which signifies stakeholders' need for forecasting models. Over time, research has been done to develop a SOTA forecasting model.

Fleischer et al. (2022) proposed a solution utilizing the LSTM neural network to learn patterns in close prices to predict the future and determine whether the closing price is enough to predict future prices. The resulting forecasts were accurate; however, the approach chosen was somewhat naïve because it was only a one-day prediction. Accurate forecasts could mean that using only the close price for forecasts is a viable base model. However, as this is an open system, it is improbable that it is sustainable.

Yenidogan et al. (2018) proposed a solution utilizing the Prophet model by Facebook to distinguish between Prophet and ARIMA. It was identified that the other available features (ex: open, volume) in the used dataset could not be used as it would yield meaningless predictions. Having removed these other features, the authors proposed adding a fiat currency price feature due to its relationship with BTC, which resulted in more accurate forecasts.

4.2.1 Factors that affect the rate of Bitcoin (BTC)

- A study conducted by Sarkodie, Ahmed and Owusu (2022) identified that the deaths from COVID-19 and the price of BTC had a very high correlation.
- Tweet sentiment - identified by Kim et al. (2016).
- Tweet volume - identified by Abraham et al. (2018) & Shen, Urquhart and Wang (2019).
- Google Trends – identified by Abraham et al. (2018).

The above-proposed solutions do not consider external factors that could impact the price of BTC. Therefore, the primary concern is that they will adapt poorly to the ever-changing and unpredictable market, where a single tweet could affect the price drastically. Although some of the factors that could influence are out of the control of researchers (ex: country legislations and laws), the factors mentioned above can cause a significant impact. It is, therefore, essential to consider these factors when building such systems.

Abraham et al. (2018) proposed a solution considering Twitter and Google Trends data. They found that the tweet volume and Google Trends were highly correlated with the price and must be included in future systems. They also identified that the tweet sentiment could be more reliable: tweets about cryptocurrencies are objective. However, they utilized a linear model for predictions, which, as identified by Maiti, Vyklyuk and Vukovic (2020), yields less performance compared to non-linear models.

3.3 Open market forecasting

As cryptocurrencies fall under the open market, it is expected to help evaluate work under other open market forecasting domains (ex: the stock market). However, what can be identified is that there needs to be more research, which signifies that it does not show promising results (Li and Bastos, 2020). However, relatively better success has been identified with the advancements in NLP and the large volumes of textual data available.

Picasso et al. (2019) proposed a solution for the stock market forecast utilizing the news data alongside historical prices. By applying sentiment analysis on the news data and combining it with the historical prices, they were able to create unison between the two schools of thought: technical analysis and fundamental analysis, which did result in a more robust solution – albeit not with an outstanding performance.

Kim et al. (2016) proposed a hybrid solution for forecasting power demand applicable to general forecasting problems. The hybrid model was implemented for ensemble potential via a bivariate learning approach. The authors also considered the external factors influencing power demand to generate a reasonable performance. It, however, is aimed at short-term power demand forecasting and is therefore limited in its application for the medium and long term.

Considering the available literature: open system forecasting is notoriously tricky, as it can depend on any external factor – forecasting the stock market is as reliable as palm reading. This is an example of one of those areas in that DL cannot be applied. Even though they are powerful algorithms that can produce performance and reliability more than humans, the obscurity of future events is a significant hindrance. It is also worth noting that the correlation between news and the stock market becomes more ambiguous when the future is considered.

4.4 Bitcoin twitter analysis

As identified above, adding external factors can improve performance drastically, where some of the more impactful factors are Twitter sentiments and tweet volumes. However, creating such models is more complex than including them as another feature in the dataset. Several issues still need solving when using tweet data as a factor in crypto price predictions (Critien, Gatt and Ellul, 2022). Some of them are detailed below:

- Bots often duplicate and automate tweets (Valencia et al., 2019).
- They are often noisy (ex: hashtags, profile mentions, URLs).
- Sarcasm can affect sentiments (Rosenthal et al., 2014).

Therefore, the data must first be preprocessed with such challenges in mind.

A point worth noting is that Valencia et al. (2019) identified that more than Twitter data is needed for the prediction of BTC, but it can be helpful when used in conjunction with other data.

Pant et al. (2018) proposed a solution using Twitter sentiment analysis for price prediction. They identified a decent correlation between Twitter sentiment and BTC's price and combined tweets' sentiment scores with historical data to predict the future price. However, as Abraham et al. (2018) identified, the tweet sentiment does not contribute as much as the tweet volume to the price of BTC, which the authors still need to include in their prediction model.

Serafini et al. (2020) explored statistical and DL models for the predictions utilizing sentiment available in the BTC ecosystem. They identified that rather than financial features, the sentiment has a more significant impact on the overall price, which further justifies that the usage of the volume and open features do not have such significance. They justified that the tweet sentiment and BTC price produced the best performance. The issue here is: although the performance obtained is excellent, the implementation is not robust since, as mentioned by Valencia et al. (2019), the Twitter data alone is not sufficient, and, as identified by Abraham et al. (2018), tweets are objective.

The drawbacks of the above two solutions are that tweets are objective, can carry sarcasm, and can be duplicated or generated by bots. Therefore, utilizing the tweet sentiment solely is not quite robust. The author believes that the optimal solution would be to utilize additional features as proposed by Abraham et al. (2018) alongside a non-linear model, as identified by Maiti, Vyklyuk and Vukovic (2020).

5. TECHNOLOGICAL REVIEW

The sheer number of available algorithmic solutions shows that this ML area still requires a satisfactory settlement. Methodologies range from classical statistical approaches to complex DL algorithms; although particular methods have proven effective, many attempts have been made to improve them. A couple of the more popular and modern techniques among the many available are evaluated below.

5.1 Statistical-based forecasting techniques

Statistical forecasting is an application of statistics to historical data to forecast what could happen in the future.

5.1.1 Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model is, in a nutshell, a combination of the Autoregressive (AR) and Moving Average (MA) models. These models predict future values based on the past (Box et al., 2015). This specific model is the most popular in TS forecasting due to its considerably good performance; however, specific vital points that cause these models to produce inaccuracies need to be considered.

- It bases its future values on its past, therefore, can be inaccurate when used in open systems.
- It is pretty complicated and hence lacks expressivity.
- It cannot be used for seasonal data.

5.1.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA is an extension to ARIMA that was introduced to handle seasonality in data. Valipour's (2015) research has proven that SARIMA produced better performance than ARIMA; however, its parameters are more sensitive to changes where even a slight change could result in poor performance.

5.1.3 Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

GARCH is an approach to estimating market volatility while being more contextual than others when predicting prices or rates (Engle, 1982). Research conducted by Bhardwaj et al. (2014) identified that the forecasts attempted by GARCH were more accurate than ARIMA's as ARIMA cannot capture volatility in the dataset.

5.1.4 Prophet

Facebook introduced an approach to solving the challenge of forecasting at scale via adjusting parameters (Taylor and Letham, 2017). It was designed to forecast daily data consisting of weekly and yearly seasonality, considering the effects of holidays (Hyndman et al., 2021). Therefore, it performs best for highly seasonal data.

Is there a clear better choice?

Upon understanding current statistical algorithms, research suggests that there is no transparent superior model as the differences in errors and accuracy are low, which signifies that the better-performing model will change according to the domain, the problem, and the dataset.

5.2 Deep learning-based forecasting techniques

Studies have shown that certain TS forecasting domains perform better on statistical models (ex: Zhang et al., 2022). At the same time, other studies have shown that DL models have outperformed statistical models (ex: Siami-Namini, Tavakoli and Siami Namin, 2018). Therefore, it is also expected to be worth evaluating DL models.

5.2.1 Long Short-term Memory (LSTM)

LSTMs were introduced by Hochreiter and Schmidhuber (1997) to solve the issues in previous RNNs of not being able to store information. They paved the path for more performant future RNNs as they do not forget short-term patterns, which made modelling temporal dependencies for larger horizons possible (Lara-Benítez, Carranza-García and Riquelme, 2021). Studies conducted to identify the impact of LSTMs on TS proved multiple advantages, the most prominent being the ability to extract meaningful information from TS data.

A prominent study among the many conducted by Sagheer and Kotb (2019) observed better performance compared to ARIMA and GRU.

5.2.2 Gated Recurrent Unit (GRU)

GRUs are, in effect, a simplification of an LSTM where the ‘forget’ and ‘input’ gates are combined. GRUs achieve similar results to LSTMs while taking less computational time (Lara-Benítez, Carranza-García and Riquelme, 2021). There is no clear distinction between GRUs and LSTMs on which one performs better; therefore, generally, both are attempted.

A study by Kuan et al. (2017) demonstrated that the GRU performance was superior to ARIMA and LSTM.

GRUs and LSTMs are effectively similar in performance

As there is a contradiction in the work of Kuan et al. (2017) and Sagheer and Kotb (2019), it is evident that there is no superior between GRU and LSTM. However, these non-linear models perform better than statistical models, in this case, ARIMA, which bestows more credibility on the study conducted by Maiti, Vyklyuk and Vukovic (2020).

5.2.3 Neural basis Expansion Analysis for interpretable Time Series (N-BEATS)

N-BEATS is a relatively new state-of-the-art forecasting algorithm. The proposed architecture aimed at solving univariate point forecasting that carries multiple properties, the most impactful being interpretable and generalizable to multiple domains (Oreshkin et al., 2020).

N-BEATS overcame the M4 competition's previous winner, a hybrid statistical and neural net model while being a pure DL architecture that is simple, generic, and expressive. This finding inspired the author to challenge the claim of Makridakis et al. (2018b): the future of TS forecasting is a hybrid of statistical and neural net models with their research.

5.2.4 Temporal Fusion Transformer (TFT)

TFT is a novel attention-based architecture proposed by Google to solve the challenge of multi-horizon forecasting optimized explicitly for outstanding performance and interpretability – as existing solutions are typically a 'black-box'. The architecture has specialized features that make it possible to choose required features and remove unnecessary ones (Lim et al., 2019). In experimentation, it was identified that there were significant performance improvements compared to existing benchmarks.

The proposed architecture was influenced upon reviewing existing transformer-based architectures (ex: Li et al., 2019) that had yet to consider different types of inputs present in multi-horizon forecasting or assumed the required exogenous features are always available.

Are DL models superior to statistical models?

Based on the above literature, it can be deduced that the belief that statistical models are superior at all times is only partially justifiable. It even cannot be deduced that non-linear DL models are superior. Therefore, it is worth investigating both schools of thought when solving TS problems and choosing the model that demonstrates better performance for that problem.

5.3 Concerns about existing used techniques

5.3.1 Issues in statistical models

Based on the above literature, the following drawbacks can be identified with statistical-based models.

- Linearity
- Suitable for seasonal data
- Lack of interpretability and expressivity
- Altering the parameters requires domain knowledge

5.3.2 Issues in deep learning models

As identified by Makridakis et al. (2018b), in contrast to other domains of NLP and computer vision, DL models still struggle in TS forecasting. Additionally, it is worth noting that although the non-linearity introduced by neural networks improves performance, they lack expressivity.

A general issue with the above models is that they are static and lack adaptability. TS data is volatile, ever-changing, and unpredictable; they can get unexpected characteristic changes to their inputs; therefore, a fixed statistical model or neural network has the possibility of struggling, which could be a reason for the identification of Makridakis et al. (2018b).

Given the open-ended conclusion on the comparison between statistical and DL models, the natural question is **how to proceed.**

5.4 Neural Ordinary Differential Equations (ODEs)

Neural ODEs are a new family of DL models that are continuous-depth, have constant memory cost, implicitly trade numerical precision for speed, and, most importantly, can adapt the evaluation strategy based on inputs (Chen et al., 2019).

Chen et al. (2019) claimed that RNNs with continuous-time hidden states determined by ODEs are effective algorithms for modelling TS data, proven by their ability to adapt computation depending on the input data. They proposed the following change to the classical equation followed by RNNs and residual networks.

Equation followed by RNNs & residual networks

$$h_{t+1} = h_t + f(h_t, \theta_t)$$

Equation proposed by Chen et al. (2019)

$$\frac{dh\,(t)}{dt} = f(h(t), t, \theta)$$

Where $h_t \in \mathbb{R}^D, t \in \{0 \dots T\}$

The equation proposed utilizes an ODE specified by a neural network to "*parameterize the derivative of the hidden units*" instead of specifying a sequence of hidden layers.

It is identified that the ability to adapt to incoming data streams could be a promising application in the domain of TS forecasting. However, these ODEs fail to identify uncertainties in predictions and are not robust (Anumasa and Srijith, 2022).

Hasani et al. (2021) mentioned upon experimentation that the performance obtained by these neural networks is underwhelming compared to a simple LSTM network. Given this fact, the author has deduced that the requirement for review of other existing neural ODEs is relatively insignificant.

5.5 Approach proposed by a Liquid Time-Constant (LTC)

Hasani et al. (2020) proposed a solution to improve the underwhelming performance of neural ODEs. The proposed architecture exhibits stable and bounded behaviour alongside being highly expressive. The alternate formulation demonstrates 'liquid' features which give rise to better performance for TS predictions.

It is also speculated that the solution can learn during training and while in use by continuously changing the underlying formulations to adapt to the changes to incoming inputs ("Liquid" machine-learning system adapts to changing conditions, 2021). What is impactful is that these networks can adapt to the changes in real-world systems while also being robust, stable, and more interpretable.

The architecture is richer in information while being scaled-down than other neural networks, which adds another advantage of it being easy to glance into the 'black box' of the underlying network to understand the reasoning behind certain computations – as realized, is another drawback of standard neural networks.

The formula is an alternative to what Funahashi and Nakamura (1993) proposed. Instead of using a neural network to define the derivative (Chen et al., 2019), an additional term assists the system in reaching equilibrium.

Equation proposed by Funahashi and Nakamura (1993)

$$\frac{dX\ (t)}{dt} = -\ \frac{X(t)}{\tau} + f(X(t), I(t), t, \theta)$$

Where $-\ \frac{\boldsymbol{X}(t)}{\tau}$ *is the additional term,* $X(t)$ *is the hidden state,* $I(t)$ *is the input, t represents time and* f *is parameterized by* θ.

Equation proposed by Hasani et al. (2020)

$$\frac{dX(t)}{dt} = -\left[\frac{1}{\tau} + \ f(X(t), \boldsymbol{I}(t), t, \theta)\right] X(t) + f(X(t), I(t), t, \theta)A$$

The above novel time-continuous RNN declares the hidden state by a system of linear ODEs parameterized by θ *and A.*

Hasani et al. (2020) further conducted experiments to justify their hypothesis producing promising results by beating other SOTA TS algorithms.

Having understood the drawbacks of the other approaches and the proposed solution by Hasani et al. (2020), the optimal way to conduct this research is by utilizing the LTC approach.

5.6 Neural Stochastic Differential Equations (SDEs)

ODEs can be used to abstract deterministic models, as utilized by Chen et al. (2019) to propose neural ODEs. Hence, the natural question arises: what could be used if there exists a state of randomness – in probabilistic models?

Neural SDEs are similar to neural ODEs in that an ODE can be used as the solver; however, the additional focus is the ‘stochastic evolution’ instead of the ‘deterministic evolution’ (Tzen and Raginsky, 2019).

Considering neural ODEs and their expressive power, SDEs can enhance the expressive power even further, as proposed by Peluchetti and Favaro (2019).

The important point worth noting about SDEs is that they are suitable for fitting data that can have an instantaneously little change (ex: market prices, molecule motion) (Duvenaud, 2021). As the chosen implementation domain belongs to this category, it is best to look into SDEs rather than ODEs going forward.

The bottom line

The usage of SDEs is more flexible as it handles randomness. The LTC architecture proposed by Hasani et al. (2020) utilizes the more obsolete ODE, which cannot model randomness and instantaneous changes. The author will therefore attempt to use the inspiration behind the LTC, but by using SDEs instead, which could manifest into a new algorithm with significant impact – which can be considered as their core contribution. The design will be presented in **chapter 4**.

5.7 Traditional required techniques

As the system would utilize multiple datasets along with the produced sentiments that would be all numerical, other more traditional techniques must also be reviewed.

5.7.1 Data preprocessing

Preprocessing is a generic step that must be performed before creating the combined enriched dataset.

5.7.1.1 Cleaning

Tweet data containing empty text fields must be ignored when calculating the average daily sentiment. Additionally, unnecessary content in text, such as hashtags, URLs, usernames, and links, must be removed (Abraham et al., 2018); however, as proposed by Hutto and Gilbert (2014), other techniques, such as punctuation and stop word removal, are best not performed. Tweets about stock trades and markets are mostly objective and have no genuine sentiment. Additionally, many of these tweets could be created by advertisements and bots, which also must be removed from the dataset (Abraham et al., 2018).

Data collected apart from tweets have a single value for each timestep. Therefore, empty rows cannot be removed, as the progression of time would be interrupted. Imputation techniques, such as 'linear interpolation', could be performed to fill in these fields as conducted by Mudassir et al. (2020).

5.7.1.2 Feature scaling

Normalization techniques are required to prevent unnecessary bias in each value. The Google Trends, historical data, Twitter volume, block reward size, and sentiments are numerical values that must be scaled to a standard range.

5.7.1.3 Feature selection

Feature selection can be conducted to select the optimal features to be an input for the model, ensuring that unneeded features with little to no impact can be ignored. Correlation is a technique that can be used to measure how much would a specific feature be of impact. Of the many available methods, the correlation matrix, 'Pearson R' – Pearson correlation coefficient, and the p-value are the primary methods of determining it (Abraham et al., 2018).

By conducting correlation tests, the author can determine whether specific features must be ignored when constructing the model, as they would not contribute as much and act as noise.

5.7.2 Hyperparameter tuning

None of the literature the author reviewed included a hyperparameter tuning stage; therefore, the author will attempt it themselves. Hyperparameter tuning is performed to get the maximum out of the model as the hyperparameters become optimal. Multiple techniques are available for tuning ML models in the scikit-learn package, such as 'grid search CV' and 'randomized search CV'. However, as this implementation will be a neural network with DL requirements, these packages cannot be used as they are computationally intensive.

Fortunately, specifically tailored techniques such as the Keras Tuner and HParams dashboard in TensorBoard, could be used to identify the best set of hyperparameters.

5.7.3 Validation

Validation techniques are required to ensure the robustness of the created model by ensuring that the predictions remain accurate in any condition. The literature reviewed by the author most utilizes the K-fold cross-validation technique (ex: Valencia et al., 2019). However, as the system is TS-based, the Walk-forward cross-validation (Serafini et al., 2020) or cross-validation on a rolling basis (Shrivastava, 2020) should be used, which are specialized versions with the ability to handle temporal data.

Conducting validation tests will ensure that the implemented model is as robust as possible, which is vital, as the chosen application domain is notorious for failures.

5.8 NLP techniques that can assist

5.8.1 Sentiment analysis

It is evident that exogenous features must be present to create a robust system. As identified by Abraham et al. (2018), NLP techniques could be utilized to extract information and obtain meaning from them alongside sentiment analysis. It is also worth noting that there may be the presence of duplicate tweets, which libraries such as fuzzywuzzy (Pant et al., 2018) could remove.

Recent research has recommended the usage of transformer architectures to apply these techniques (Wolf et al., 2020). However, the author believes that to extract the sentiment from the data, the Vader library is superior, as it can effectively extract sentiment in the context of social media (Hutto and Gilbert, 2014), and it is furthermore more tailored for Twitter data (Valencia et al. 2019). Nevertheless, both will be evaluated.

It is worth mentioning that specific cleaning steps must not be performed. These steps include the removal of punctuation, capitalization, and removal of stop words – especially ‘but’. Hutto and Gilbert (2014) developed five heuristics that are utilized within Vader; the most prominent four are described below:

- Punctuation – can increase the magnitude of intensity without modifying the sentiment.
- Capitalization – acts in a similar way to punctuation.
- ‘But’ – could signal a shift in polarity.
- Degree modifiers – act in a similar way to punctuation and capitalization.

5.8.2 Named Entity Recognition (NER)

Upon conducting sentiment analysis, NER can be performed to adjust the weight (impact) of each tweet’s sentiment depending on how influential the tweeter is or any organization being mentioned (ex: a tweet of Elon Musk could be given more weightage as a single tweet from him could cause a drastic impact).

5.9 Proposed architecture

Upon identifying the requirements to implement the system and considering required exogenous factors, the author proposes the following architecture to fulfil the research's mentioned aim.

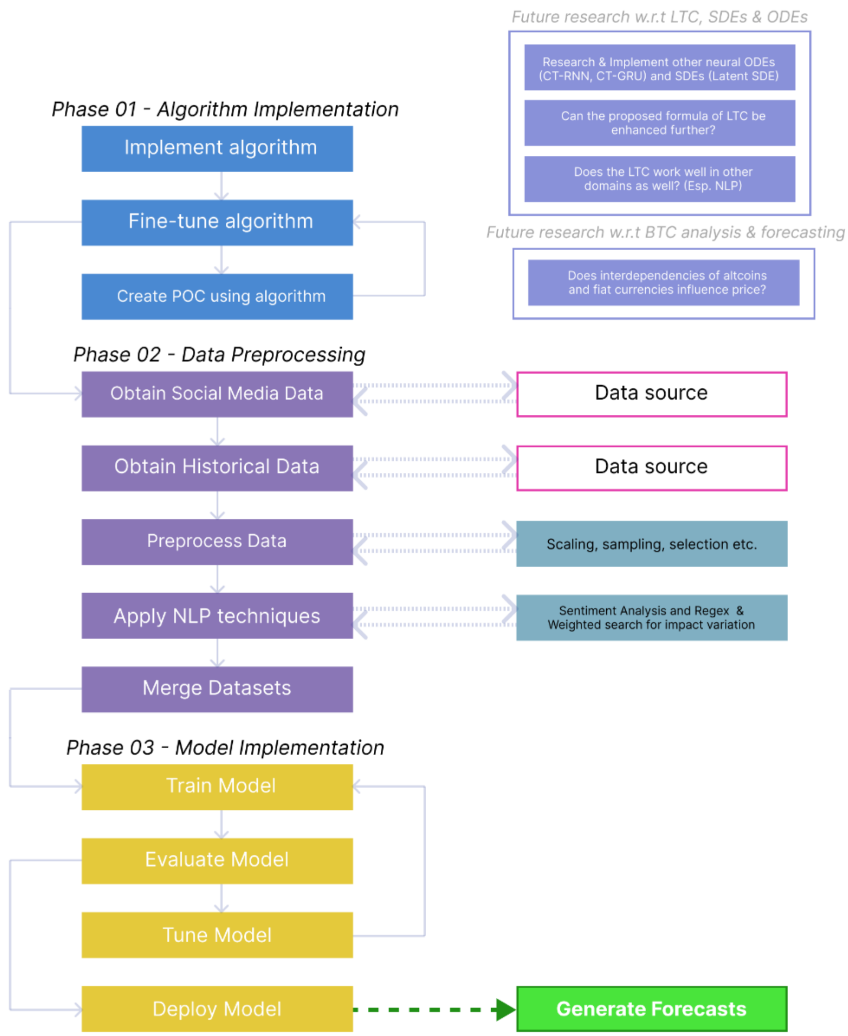

Figure 3: Proposed Architecture (*Self-Composed*)

As identified, the architecture will use social media data to make the forecasting model as robust as possible; the data would be fetched from an API which could be Google Trends, Twitter tweets, and the block reward size.

6. EVALUATION

6.1 Evaluation approaches

Evaluation of the application can be conducted by validation methodologies and evaluation metrics based on generating forecasts and comparing them against the test data.

The evaluation metrics: MAE, MSE, RMSE, and MAPE (Hyndman et al., 2021) can deduce how well the model performs, where the lower the value obtained, the better the performance. It is worth noting that it is implausible to get values of 0 – obtaining such errors is a clear distinction that something has gone wrong.

Table 1: Evaluation Metrics for TS forecasting systems

Metric	Description	Formulae
MAE	The absolute difference between the generated forecast and the ground truths. Since it is simple to understand, it will give the author an idea of how inaccurate the forecast is.	$\frac{\sum_{i=1}^{n} \lvert y_i - \hat{y_i} \rvert}{n}$
MSE	Similar to MAE but computes the squared differences. This metric gives more emphasis to more significant errors and also considers outliers.	$\frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{n}$
RMSE	The square root of MSE.	$\sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{n}}$
MAPE	The ratio of the forecast to the average absolute difference between the forecast and ground truths. Helpful if the algorithm is evaluated in the M4 competition, as this metric is used widely in competitions since it does not have any units. Therefore, it can be compared across other datasets.	$\frac{100}{n} \sum_{i=1}^{n} \lvert \frac{y_i - \hat{y_i}}{y_i} \rvert$

An additional metric: MASE, can be used, which is an alternative to percentage errors (ex: MAPE) proposed by Hyndman & Koehler (2006) as there is a possibility for MAPE to 'explode'. Helpful when trying to beat the naïve forecast, which produces a MASE value close to 1; therefore, getting a value less than this would mean that the model performs better than the naïve forecast.

Apart from these technical metrics, quality-of-service could be validated, such as CPU & memory usage, startup time, and application fluidity.

6.2 Benchmarking

6.2.1 System benchmarking

A standard dataset is required to conduct a valid benchmarking analysis on the system. As there is no previous system utilizing the LTC and no system combining all mentioned features in a non-linear model, the author will not be able to conduct a benchmarking analysis on the proposed system. However, conducting baselining to capture performance is feasible.

Moreover, the results obtained will be publicly available to aid future research and benchmarking analysis.

6.2.2 Algorithm benchmarking

The algorithm could be benchmarked in the M4 competition, as specific benchmarking datasets exist.

6.3 Hypothesis justification

To wholly justify the author's hypothesis, two factors of evaluation must take place:

- **E**01 - evaluate the application against existing work of cryptocurrency forecasting.
- **B**01 - benchmark algorithm against other SOTA TS algorithms.

As identified, LTC applications do not exist that could be compared. Therefore, the author will compare against the existing work done in forecasting cryptocurrencies to fulfil **E**01. Additionally, if time permits, the author will attempt to fulfil **B**01 by benchmarking the algorithm in the M4 competition.

REFERENCES

Fleischer, J.P. et al. (2022). Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory. *Algorithms*, 15 (7), 230. Available from https://doi.org/10.3390/a15070230 [Accessed 26 September 2022].

Maiti, M., Vyklyuk, Y. and Vuković, D. (2020). Cryptocurrencies chaotic co-movement forecasting with neural networks. *Internet Technology Letters*, 3 (3). Available from https://doi.org/10.1002/itl2.157 [Accessed 16 October 2022].

Yenidogan, I. et al. (2018). Bitcoin Forecasting Using ARIMA and PROPHET. *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. September 2018. Sarajevo: IEEE, 621–624. Available from https://doi.org/10.1109/UBMK.2018.8566476 [Accessed 16 October 2022].

Abraham, J., Higdon, D., Nelson, J. and Ibarra, J. (2018). Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review:* Vol. 1: No. 3, Article 1. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/1

Shen, D., Urquhart, A. and Wang, P. (2019). Does twitter predict Bitcoin? *Economics Letters*, 174, 118–122. Available from https://doi.org/10.1016/j.econlet.2018.11.007 [Accessed 16 October 2022].

Kim, Y.B. et al. (2016). Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. *PLOS ONE*, 11 (8), e0161197. Available from https://doi.org/10.1371/journal.pone.0161197 [Accessed 16 October 2022].

Sarkodie, S.A., Ahmed, M.Y. and Owusu, P.A. (2022). COVID-19 pandemic improves market signals of cryptocurrencies–evidence from Bitcoin, Bitcoin Cash, Ethereum, and Litecoin. *Finance Research Letters*, 44, 102049. Available from https://doi.org/10.1016/j.frl.2021.102049 [Accessed 16 October 2022].

Li, A.W. and Bastos, G.S. (2020). Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review. *IEEE Access*, 8, 185232–185242. Available from https://doi.org/10.1109/ACCESS.2020.3030226 [Accessed 16 October 2022].

Picasso, A. et al. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, 135, 60–70. Available from https://doi.org/10.1016/j.eswa.2019.06.014 [Accessed 16 October 2022].

Kim, M. et al. (2019). A Hybrid Neural Network Model for Power Demand Forecasting. *Energies*, 12 (5), 931. Available from https://doi.org/10.3390/en12050931 [Accessed 16 October 2022].

Critien, J.V., Gatt, A. and Ellul, J. (2022). Bitcoin price change and trend prediction through twitter sentiment and data volume. *Financial Innovation*, 8 (1), 45. Available from https://doi.org/10.1186/s40854-022-00352-7 [Accessed 16 October 2022].

Valencia, F., Gómez-Espinosa, A. and Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy*, 21 (6), 589. Available from https://doi.org/10.3390/e21060589 [Accessed 16 October 2022].

Rosenthal, S. et al. (2014). SemEval-2014 Task 9: Sentiment Analysis in Twitter. Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014. *Dublin, Ireland: Association for Computational Linguistics*, 73–80. Available from https://doi.org/10.3115/v1/S14-2009 [Accessed 16 October 2022].

Pant, D.R. et al. (2018). Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. 2018 IEEE 3rd International Conference on Computing, *Communication and Security (ICCCS)*. October 2018. Kathmandu: IEEE, 128–132. Available from https://doi.org/10.1109/CCCS.2018.8586824 [Accessed 16 October 2022].

Serafini, G. et al. (2020). Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches. *2020 International Joint Conference on Neural Networks (IJCNN)*. July 2020. Glasgow, United Kingdom: IEEE, 1–8. Available from https://doi.org/10.1109/IJCNN48605.2020.9206704 [Accessed 16 October 2022].

Lara-Benítez, P., Carranza-García, M. and Riquelme, J.C. (2021). An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems*, 31 (03), 2130001. Available from https://doi.org/10.1142/S0129065721300011 [Accessed 16 October 2022].

G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time series analysis: forecasting and control (John Wiley & Sons, 2015).

Valipour, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States: Runoff forecasting using SARIMA. *Meteorological Applications*, 22 (3), 592–598. Available from https://doi.org/10.1002/met.1491 [Accessed 16 October 2022].

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50 (4), 987. Available from https://doi.org/10.2307/1912773 [Accessed 28 December 2022].

Bhardwaj, S.P. et al. (2014). An Empirical Investigation of Arima and Garch Models in Agricultural Price Forecasting. *Economic Affairs*, 59 (3), 415. Available from https://doi.org/10.5958/0976-4666.2014.00009.6 [Accessed 17 October 2022].

Taylor, S.J. and Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*. Available from https://doi.org/10.7287/peerj.preprints.3190v2 [Accessed 17 October 2022].

Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. Available from https://otexts.com/fpp3/. [Accessed on 30 Sep. 2022].

Zhang, R. et al. (2022). Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. *PLOS ONE*, 17 (1), e0262009. Available from https://doi.org/10.1371/journal.pone.0262009 [Accessed 17 October 2022].

Siami-Namini, S., Tavakoli, N. and Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. December 2018. Orlando, FL: IEEE, 1394–1401. Available from https://doi.org/10.1109/ICMLA.2018.00227 [Accessed 17 October 2022].

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9 (8), 1735–1780. Available from https://doi.org/10.1162/neco.1997.9.8.1735 [Accessed 25 September 2022].

Sagheer, A. and Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323, 203–213. Available from https://doi.org/10.1016/j.neucom.2018.09.082 [Accessed 17 October 2022].

Kuan, L. et al. (2017). Short-term electricity load forecasting method based on multilayered self-normalizing GRU network. *2017 IEEE Conference on Energy Internet and Energy System Integration* (EI2). November 2017. Beijing: IEEE, 1–5. Available from https://doi.org/10.1109/EI2.2017.8245330 [Accessed 17 October 2022].

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018b). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34 (4), 802–808. Available from https://doi.org/10.1016/j.ijforecast.2018.06.001 [Accessed 25 September 2022].

Oreshkin, B.N. et al. (2020). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. Available from http://arxiv.org/abs/1905.10437 [Accessed 26 September 2022].

Lim, B. et al. (2019). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. Available from https://doi.org/10.48550/ARXIV.1912.09363 [Accessed 17 October 2022].

Li, S. et al. (2019). Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. Available from https://doi.org/10.48550/ARXIV.1907.00235 [Accessed 17 October 2022].

Chen, R.T.Q. et al. (2019). Neural Ordinary Differential Equations. Available from https://doi.org/10.48550/arXiv.1806.07366 [Accessed 25 September 2022].

Hasani, R. et al. (2021). Liquid Neural Networks. *YouTube*. Available from https://www.youtube.com/watch?v=IlliqYiRhMU&t=350s. [Accessed on 30 Sep. 2022].

Hasani, R. et al. (2020). Liquid Time-constant Networks. Available from https://doi.org/10.48550/arXiv.2006.04439 [Accessed 25 September 2022].

Anumasa, S. and Srijith, P.K. (2022). Latent Time Neural Ordinary Differential Equations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36 (6), 6010–6018. Available from https://doi.org/10.1609/aaai.v36i6.20547 [Accessed 17 October 2022].

"Liquid" machine-learning system adapts to changing conditions. (2021). *MIT News | Massachusetts Institute of Technology*. Available from https://news.mit.edu/2021/machine-learning-adapts-0128 [Accessed 17 October 2022].

Funahashi, K. and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6 (6), 801–806. Available from https://doi.org/10.1016/S0893-6080(05)80125-X [Accessed 14 October 2022].

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225

Wolf, T. et al. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. October 2020. Online: Association for Computational Linguistics, 38–45. Available from https://doi.org/10.18653/v1/2020.emnlp-demos.6 [Accessed 19 October 2022].

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688

Shrivastava, S. (2020). Cross Validation in Time Series. *Medium*. Available from https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4 [Accessed 12 October 2022].

Ahmed, NK., Atiya, AF., Gayar, NE. and El-Shishiny, H. An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*. 2010;29(5-6):594–621.

Lim, B. and Zohren, S. (2020). Time Series Forecasting With Deep Learning: A Survey. Available from https://doi.org/10.1098/rsta.2020.0209 [Accessed 21 October 2022].

S. Nakamoto, (2020). *Bitcoin: A peer-to-peer electronic cash system*. Available from https://bitcoin.org/bitcoin.pdf [Accessed 25 September 2022].

Kfir, I. (2020). Cryptocurrencies, national security, crime and terrorism. *Comparative Strategy*, 39 (2), 113–127. Available from https://doi.org/10.1080/01495933.2020.1718983 [Accessed 22 October 2022].

Bouri, E., Gupta, R., Lau, C. K. M., Roubaud, D., & Wang, S. (2018). Bitcoin and global financial stress: A copula-based approach to dependence and causality in the quantiles. *The Quarterly Review of Economics and Finance*, 69, 297–307.

Rejeb, A., Rejeb, K. and G. Keogh, J. (2021). Cryptocurrencies in Modern Finance: A Literature Review. *ETIKONOMI*, 20 (1), 93–118. Available from https://doi.org/10.15408/etk.v20i1.16911 [Accessed 22 October 2022].

Wilson, C. (2019). Cryptocurrencies: The Future of Finance? In: Yu, F.-L.T. and Kwan, D.S. (eds.). *Contemporary Issues in International Political Economy*. Singapore: Springer Singapore, 359–394. Available from https://doi.org/10.1007/978-981-13-6462-4_16 [Accessed 22 October 2022].

Buhalis, D. et al. (2019). Technological disruptions in services: lessons from tourism and hospitality. *Journal of Service Management*, 30 (4), 484–506. Available from https://doi.org/10.1108/JOSM-12-2018-0398 [Accessed 22 October 2022].

Wang, Y. et al. (2019). Making sense of blockchain technology: How will it transform supply chains? *International Journal of Production Economics*, 211, 221–236. Available from https://doi.org/10.1016/j.ijpe.2019.02.002 [Accessed 22 October 2022].

Pournader, M. et al. (2020). Blockchain applications in supply chains, transport and logistics: a systematic review of the literature. *International Journal of Production Research*, 58 (7), 2063–2081. Available from https://doi.org/10.1080/00207543.2019.1650976 [Accessed 22 October 2022].

Nica, O., Piotrowska, K., & Schenk-Hoppé, K. R. (2017). Cryptocurrencies: Economic Benefits and Risks. SSRN Scholarly Paper ID 3059856.

Alonso-Monsalve, S. et al. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, 149, 113250. Available from https://doi.org/10.1016/j.eswa.2020.113250 [Accessed 22 October 2022].

Cryptocurrency Market Size, Growth & Trends | Forecast [2028]. (2021). Available from https://www.fortunebusinessinsights.com/industry-reports/cryptocurrency-market-100149 [Accessed 23 October 2022].

Avdeychik, V., & Capozzi, J. (2018). SEC's Division of Investment Management Voices Concerns Over Registered Funds Investing in Cryptocurrencies and Cryptocurrency Related Products. *Journal of Investment Compliance*, 19(2), 8–12.

Baldimtsi, F., Kiayias, A., & Samari, K. (2017). Watermarking Public-Key Cryptographic Functionalities and Implementations. In Nguyen, P. Q., & Zhou, J. (Eds.). *Information Security*, 173–191. Berlin: Springer.

Kerr, J. (2018). How Can Legislators Protect Sport from the Integrity Threat Posed by Cryptocurrencies? *The International Sports Law Journal*, 18(1), 79–97.

Miller, P. (2016). Chapter 1—The Cryptocurrency Enigma. In Sammons, J. (Ed.), *Digital Forensics*, 1–25. Syngress. https://doi.org/10.1016/B978-0-12-804526-8.00001-0

Scharding, T. (2019). National Currency, World Currency, Cryptocurrency: A Fichtean Approach to the Ethics of Bitcoin. *Business and Society Review*, 124(2), 219–238.

Want to Buy Crypto? Here's What to Look for In a Crypto Exchange. (2021). *Time*, 11 June. Available from https://time.com/nextadvisor/investing/cryptocurrency/what-are-cryptocurrency-exchanges [Accessed 23 October 2022].

Peluchetti, S. and Favaro, S. (2019). Infinitely deep neural networks as diffusion processes. Available from https://doi.org/10.48550/ARXIV.1905.11065 [Accessed 2 December 2022].

Tzen, B. and Raginsky, M. (2019). Neural Stochastic Differential Equations: Deep Latent Gaussian Models in the Diffusion Limit. Available from http://arxiv.org/abs/1905.09883 [Accessed 2 December 2022].

Duvenaud, D (2021). Directions in ML: Latent Stochastic Differential Equations: An Unexplored Model Class. *YouTube*. Available from https://www.youtube.com/watch?v=6iEjF08xgBg. [Accessed on 30 Sep. 2022].

Mudassir, M. et al. (2020). Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. *Neural Computing and Applications*. Available from https://doi.org/10.1007/s00521-020-05129-6 [Accessed 3 December 2022].

APPENDIX A – CONCEPT MAP