INTERPRETABLE MULTI-HORIZON TIME SERIES FORECASTING OF CRYPTOCURRENCIES BY LEVERAGING TEMPORAL FUSION TRANSFORMERS

by

Arslan Farooq

(DS320192002)

Supervisor-I	Dr. M. Irfan Uddin	Institute of Computing, KUST, Kohat	.... Signature
Supervisor-II	Dr. Muhammad Adnan	Institute of Computing, KUST, Kohat	.... Signature
Director IOC	Prof. Dr. Shafiullah Khan	Institute of Computing, KUST, Kohat	.... Signature

Institute of Computing
Kohat University of Science & Technology, Kohat-26000,
Khyber Pakhtunkhwa, Pakistan.

Introduction

The majority of current architectures for predicting cryptocurrency prices are 'black-box' models, where forecasts are controlled by complex nonlinear interactions between many parameters. The time ordering of the input features is not considered by traditional methods in their conventional form, such as LIME and SHAP [1]. As relationships between time steps are frequently significant in time series, such frameworks would result in poor interpretation quality. So, that makes most times less explainable which is a problem because, when applied, it is good to know why certain decisions are being made and why specific predictions start happening.

This research aims to use the proposed model for cryptocurrency markets to forecast significant events and make better-informed investment decisions. The Temporal Fusion Transformers (TFT) model comprises different components capable of attention analysis of time steps and also telling how much each feature contributes to forecasting. However, the performance and accuracy of TFT are greatly influenced by their hyperparameters. Choosing the best collections of six hyperparameters is a challenging task. The Adaptive Differential Evolution (ADE) method will increase forecast accuracy and stability by assigning optimized hyperparameters to the Temporal Fusion Transformers model. Which will produce an explicable time-dynamic interpretable analysis, and the model is named ADE-TFT.

Reddy et al. compare the LASSO machine-learning algorithm with other models like QUANDL, RNN, SVM, and CNN. LASSO offers excellent time management, allowing to achieve superior results from a large dataset. They have the highest level of accuracy in predicting daily trending changes in the bitcoin market [2]. Derbentsev forecasted the bitcoin values by using two machine learning algorithms, stochastic gradient boosting machine (SGBM) and random forest (RF). Results demonstrate that machine learning approaches can estimate the bitcoin values. However, the decisions were made at the appropriate time to reduce the risks associated with the investment decision [3]. Moreover, Kumar et al. demonstrated that everyday data comprised 1000 data samples, hourly data of 1500 data samples, and minute data of 400000 data lines. They discovered how DL algorithms enabled the pricing trends in Ethereum cryptocurrencies. For forecasting Ethereum value and map Mean Absolute Percentage (MAPE) errors, authors used either Multi-Layer Perceptron (MLP) or Long Short-Term Memory (LSTM) models [4].

Lopez and other scholars collected data from six different facilities in Germany and Australia by applying the TFT model to predict hourly day-ahead PV power generation with statistical error indicators to compare the outcomes with other models. TFT has shown more precise results than the other algorithm to forecast PV power generation [5]. Furthermore, Feng tackles the supply air temperature in high-speed train carriages, by enhancing two TFT architecture components: the Double-Convolutional Residual Encoder and the spatio-temporal Double-Gated. Additionally, a loss function is also developed which is appropriate for generic long-sequence time-series forecast tasks for predicting temperature, that increases MAPE by 11.73% and 21.70% compared to the original model [6].

Research Problem

Prices of cryptocurrencies depend on new technologies, market pressure to deliver, usage and costs, security threats, and cultural conditions. Price prediction is a challenging task due to the lack of indicators; In comparison to traditional financial predictions, such as stock market predictions, cryptocurrencies are relatively unpredictable. Investing in digital currencies involves more risk and less profit. This research uses a novel model named ADE-TFT along with fine-tuning hyperparameters where the fine-tuning is based on Adaptive Differential Evolution, to predict the quantiles of prices with a low error rate and high accuracy precision. Deep learning architecture techniques can trend indicators control the cryptocurrency's price hike. The model proposed in this research will not tell the exact future price values. However, the proposed model will forecast the price quantiles dependent on known inputs, observed inputs, and static covariates from historical data with a clear direction to expect the prices to move.

Objectives

- To determine how precise the ADE-TFT model will forecast future results.
- To show the Variable importance in forecasting by using interpret model.
- To evaluate model performance using supervised metrics techniques.

Research Methodology

Temporal Fusion Transformers
Google built an attention-based deep neural network architecture to integrate multi-horizon forecasting with clear perception into temporal dynamics. An attention mechanism adds new levels of interpretability to the model. Ranking input variables based on the size of their attitude weights is possible, which enables us to assess the relative significance of the inputs and find nonperforming variables. Eliminating underperforming variables improves prediction accuracy and simplifies computation [1].
Other key features of the Temporal Fusion Transformer are:
a) Gating Residual Network (GRN) that supports multiple datasets and dilemmas by dropping the unnecessary elements of architecture to avoid nonlinear processing. b) Variable Selection Network (VSN), at each time step, picks out appropriate input variables. c) Static covariates encoders that give out context vectors by capturing the critical characteristics of sequence input vectors. d) Temporal processing, by taking known and observed input, learns both time series patterns, whether long-term or short-term. e) Prediction intervals, which predict the future values in quantiles (the median values between 10-90 percent quantiles) rather than predicting specific values.
Proposed ADE-TFT Model
After manipulation, data is divided into three sets test, training, and validation. In ADE, the Mutant Operator creates a mutant/donor vector from the parent vector that will be selected randomly in each iteration from the total population

and then passes those mutant/donor vectors to the Crossover Operation, in which all the elements of the target vector are combined with the donor/mutant vector with the probability Crossover Range (CR) and creates trial vector offspring. After this Tournament Selection method is used to select the trial vector, and the fitness value compare with the target vector using the fitness proportional selection function [7],

$$\mathrm{P_i} = \frac{f(x_i)}{\sum_{i=0}^{\mu-1} f(x_i)} \tag{1}$$

the one with the best fitness value replaces with the target vector. This process continued until all six hyperparameters of the TFT model were optimized. The optimized hyperparameters are learning rates, batch sizes, the number of time steps, hidden layer counts, consecutive hidden layer counts, and attention head counts. These optimized hyperparameters will pass to our TFT model for the prediction intervals. Figure 1 illustrates the flow of the proposed ADE-TFT model.

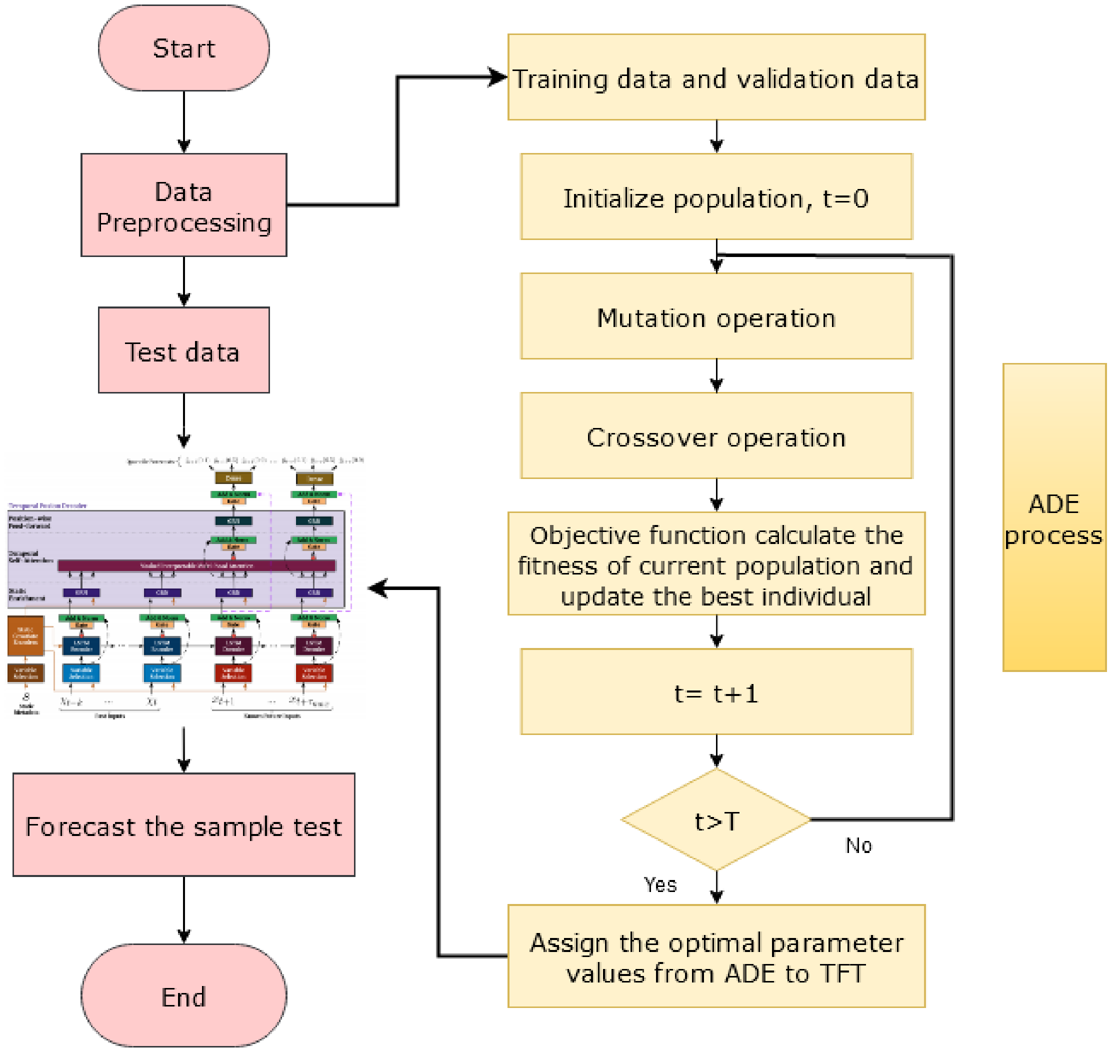

Figure 1: Flowchart of proposed ADE-TFT model

Forecasting Framework of this Research

The most crucial phase of our research, based on the entire technique, is choosing an appropriate dataset with historical data for Bitcoin cryptocurrency. A variety of Bitcoin cryptocurrency datasets have been examined and the following primary features will be extracted using an API and stored in CSV. Those features are Market_cap, volume, crypto_name, low, close, open, high, time, price_usd, price_btc, adj_close, date, principle_market_price_usd,principle_price_usd. The raw data file contains denormalized or poorly structured data and irrelevant features. Python packages will be implemented to handle that type of data: NumPy to make data-frames for fundamental scientific computations, Pandas for data manipulation and feature selection using statistical tools, Matplotlib, and Seaborn for data visualization and exploratory data analysis. Processed data then pass to proposed algorithm which produces precise forecasts by considering all relevant variables and previous patterns in the data. We have specified that short-term forecasting may range from one day to a week, while long-term forecasting may range from a week to month.

Our learning algorithm, ADE-TFT, can effectively capture and analyze these patterns to adjust the forecasting model and provide more accurate predictions during a recession period. By incorporating variables that capture the indicators of a potential recession into the model, our algorithm can adjust the forecasted prices and provide more reliable predictions, even during periods of economic decline. This information is critical to provide clarity on the timeframe of our forecasting models and ensure that our predictions are suitable for different time horizons. Model performance evaluation will be conducted to test how much each variable contributes to predicting future prices, as shown in Figure 2.

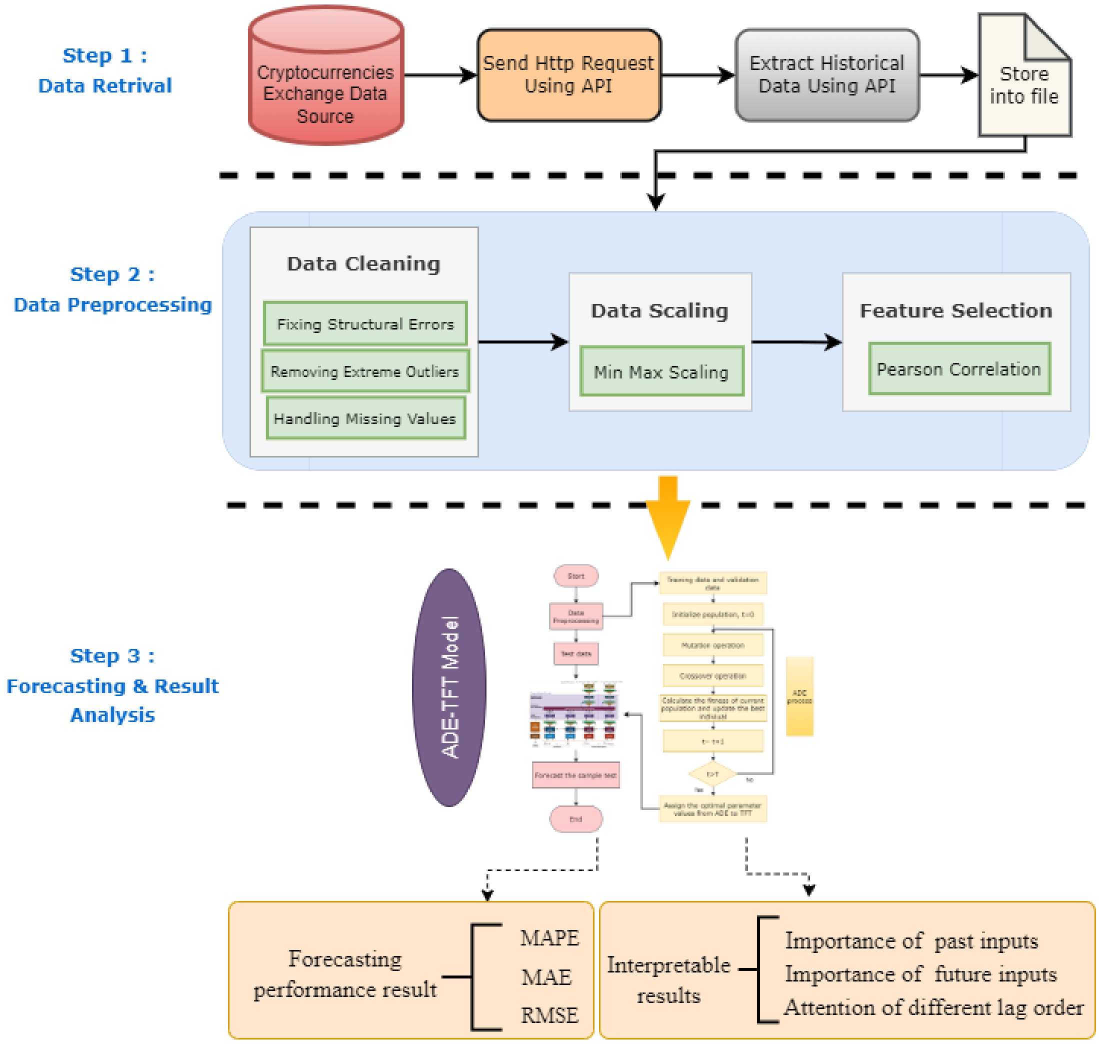

Figure 2: Workflow diagram for time series forecasting

Time Frame

S.No.	Research Components	Proposed Time
1	Literature Review	03 months
2	Initial design and data collection	02 months
3	System implementation and testing	01 months
4	System Evaluation	02 weeks
5	Thesis writing	03 months

References

[1] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[2] L. S. Reddy and P. Sriramya, "A research on bitcoin price prediction using machine learning algorithms," *Int. J. Sci. Technol. Res*, vol. 9, no. 4, pp. 1600–1604, 2020.

[3] V. Derbentsev, V. Babenko, K. Khrustalev, H. Obruch, and S. Khrustalova, "Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices," *International Journal of Engineering*, vol. 34, no. 1, pp. 140–148, 2021.

[4] D. Kumar and S. Rath, "Predicting the trends of price for ethereum using deep learning techniques," in *Artificial Intelligence and Evolutionary Computations in Engineering Systems*. Springer, 2020, pp. 103–114.

[5] M. López Santos, X. García-Santiago, F. Echevarría Camarero, G. Blázquez Gil, and P. Carrasco Ortega, "Application of temporal fusion transformer for day-ahead pv power forecasting," *Energies*, vol. 15, no. 14, p. 5232, 2022.

[6] G. Feng, L. Zhang, F. Ai, Y. Zhang, and Y. Hou, "An improved temporal fusion transformers model for predicting supply air temperature in high-speed railway carriages," *Entropy*, vol. 24, no. 8, p. 1111, 2022.

[7] A. M. Andrew, "Introduction to evolutionary computing, by ae eiben and je smith (natural computing series), springer, berlin, 2003, hardback, xv+ 299 pp., isbn 3-540-40184-9 (£ 30.00)," *Robotica*, vol. 22, no. 3, pp. 349–349, 2004.