Predicting Structured and Volatile Time Series: A Machine Learning Approach to Electrical Power Consumption and Bitcoin Prices

MSc Research Project
MSc. Financial Technology

JISNO GEORGE
Student ID: 23324937

School of Computing
National College of Ireland

Supervisor: Brian Byrne

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……. ………………Jisno George ……………………………………………………		
Student ID:	……………………………23324937…………………………………………………………..…….		
Programme:	…MSc in FinTech………………………………	**Year:**	2024-25
Module:	…………………………………………………………………………………………………….………		
Supervisor:	……………Brian Byrne………………………………………………………………..………		
Submission Due Date:	……………11-08-2025…………………………………………………………..………		
Project Title:	Predicting Structured and Volatile Time Series: A Machine Learning Approach to Electrical Power Consumption and Bitcoin Prices		
Word Count:	………5667………………… **Page Count**……………………23…………………….……..		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………………………Jisno GEORGE………………………………………………

Date: ……………………11-08-2025…………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting Structured and Volatile Time Series: A Machine Learning Approach to Electrical Power Consumption and Bitcoin Prices

JISNO GEORGE
23324937

1. Abstract

The paper examines how machine learning (ML) regression can be utilized, namely XGBoost, to predict the short-term residential electricity use based on a daily aggregated electricity consumption data with respect to 2017. The temporal patterns were reflected in covariates engineered in the dataset, including lagged consumption values, rolling averages, and day-of-week (e.g., variables). XGBoost had an MAPE rate of around 6.89 percent, and RMSE of 0.415 kWh revealing top-notch predictive accuracy and approximation with real-world values of consumption. The results of this performance highlight the suitability of tree-based ensemble to structured, seasonal data. To set a comparative background, the study replicated the Bitcoin price data, which is a non-stationary and volatile time series (20112021) using the approach of modeling. In this case, the predictive accuracy was lower (RMSE in the hundreds of USD), as is the case when market volatility, poor seasonality, and reliance on exogenous determinants such as regulatory change or market mood make it difficult to model. This comparison shows the significance of dataset properties in the establishment of performance of the model. Tree-based models perform well in stable and seasonal data such as power consumption but not in volatile data, which need an adapting or the hybrid model that includes external interventions.

Expanded Insights:

- **Significance**: Such a high accuracy of power consumption forecasting enables useful applications in operations, such as load balancing and renewable energy integration, which are very relevant in modern energy systems (known to become more electrified in the future e.g., electric vehicles), that are under pressure due to climate-driven changes in demand.
- **Bitcoin Comparison**: Lower accuracy in Bitcoin forecasting (compared with FX sales in the earlier example) is an indicator of the importance of exogenous data (e.g., social media sentiment, macroeconomic indicators) to be used to model the market dynamics, which cannot be reliably modeled by using only historical prices.

- **Practical Implications**: The results imply that energy providers can utilise ML models to manage the demand cost-sufficiently, whereas financial analysts require hybrid models to deal with the crypto markets.
- **Future Directions**: An extension of the study to 2025 statistics may include renewable energy penetration dynamics, streaming data analysis, or more sophisticated routines such as Transformers to operate better in the two fields.

2.Introduction

The short-term prediction of electrical power usage is an important aspect of reliable energy generation, conveyance and demand-side practice. As renewable energy sources become more fully integrated, as transport becomes increasingly electrified, and as the consumption behaviour changes due to climate factors, the potential to forecast demand with a high level of precision helps to contribute to economic cost-effectiveness as well as sustainability. Particularly tree-based ensembles and recurrent neural networks have shown to be promising towards a regression-based machine learning (ML) paradigm, capable of capturing the non-linear seasonal pattern that residential energy use often follows. This study explores and assesses the use of machine learning regression models, especially the XGBoost algorithm, on forecasting the daily electrical consumption based on previous consumption data and also using engineered time-based features. In an effort to make this generalization, the research compares these findings to the behavior of other comparable modelling frameworks on a fluctuating, non-stationary example within the Bitcoin market. This comparative study has two uses:

1. The reason is to demonstrate the role of statistical characteristics of a dataset on the performance of a model.
2. To put a perspective on how model choice should coincide with data structure and stability of the target data.

The reasons behind such a study are the need to fill the gap between the theoretical advances in predictive modelling and its practical adaptation in work. Accurate forecasts can be used by energy providers to allow proactive load balancing, enhanced integration of intermittent renewable generation and affordable capacity planning. In financial markets, predictive accuracy is limited according to the volatility and external events inherently, both model performance insights and hybrid methods involving exogenous signals can be informed.

This work is being driven by the question:

Are machine learning regression models capable of predicting short-term estimates of power consumed by homes and, what is their performance when used on datasets with very different statistical characteristics?

The issue of this question is reflected in the research that provides a degree of methodological clarity, empirical performance points of reference and practical prescriptions of interest to both academic researchers and industry practitioners. The implications to the expected design of sound forecasting systems, flexible to the peculiarities of separate application areas, are defined in the findings.

3. Literature Review

The overlap between the latest modelling methods of energy consumption in the power systems and the prediction of cryptocurrency price is linked to the critical issues of resources optimization, risk management, and sustainability.The guiding research question that will guide is: *How can machine learning, deep learning, and optimisation frameworks model complex dynamic systems in a way that improves predictive accuracy while promoting sustainable outcomes?*

This review summarises 25 peer-reviewed IEEE articles (20162025) on applications in energy grids, data centres, trading markets of cryptocurrencies. The thematic organisation is comprised of four main frameworks namely multi-objective optimisation, predictive modelling, component-level and empirical modelling as well as hybrid frameworks that have exogenous data integration. Both sections analyze perceived strengths and weaknesses of the available methods with regards to how well they handle nonlinearities and temporal dependencies and how easy they adapt to real-time volatility.

I. Multi-Objective Optimisation for Resource Allocation

The use of multi-objective optimisation is essential in energy systems, in which trade-offs need to be balanced between cost, reliability, risk, and sustainability, as well as in cryptocurrency markets, in which trade-offs are non-existent between cost, reliability, risk, and sustainability. Hamidizadeh et al. [5] created a Pareto-optimal multi-objective model of the National Iranian Oil Products Distribution Company that decreased energy consumption without violating supply chain and attained efficiencies of up to 15 percent. Jiang et al. [7] presented two levels of model inter-provincial power purchase that places priorities on the integration of renewable energy to improve stability in the grid and reductive energy losses in the transmissions. This was generalised by Yue et al. [22] to hierarchical optimisation of high-voltage DC grids, with the load and efficiency of regions being balanced. Toleti et al. [21] applied Hidden Markov Models to combine with ML-based trading systems in cultivating adaptations to volatility-related market regime changes that boosted Sharpe ratios by determining state transitions in volatile environments in cryptocurrency domain. Lv et al. [9] showed solar power plant optimisation considering voltage stability on the grid, the curtailment minimisation, and clean energy uptake, an idea that is conceptually similar to risk-return optimisation in finance systems. These frameworks are best where there is control, but they are vulnerable to stochastic phenomena like renewable intermittency or sentiment based price spikes because they rely too much on deterministic assumptions. Further developments may add a stochastic programming or deep reinforcement learning in order to be more adaptive in the turbulent conditions.

II. Predictive Modelling for Dynamic Systems

Cryptocurrency price prediction as well as energy demand prediction still revolves around predictive modelling. Wang et al. [23] used Vector Error Correction (VEC) models to predict the provincial electricity demand in China with an error rate of less than 10 per cent; they attributed this to the consumption behaviour and match it with macroeconomic variables like GDP. Random Forests were used by Zhong et al. [25] to forecast energy consumption on the servers, achieving RMSE that is less than 5 percent to capture the correlation between the data and non-linear patterns. Tian et al. [20] used LSTM networks, which were applied to model data centre cooling power, and that make use of the time sequence to provide accurate predictions in the short term. Periketi et al. [16] and Ramani et al. [17] also showed that LSTM with Prophet also decreased MAPE by up to 30 per cent relative to ARIMA in cryptocurrency markets to model the candlestick pattern. The study by Pattanayak et al.

[15] investigated the GRU models which provided efficiencies at no loss of accuracy. According to Sahoo et al. [18], the comparison between parametric and non-parametric approaches showed that ARIMA-type models are reliable with short horizons, however, they are ineffective in high levels of volatility. Mahfooz and Phillips [12] demonstrated that this expanded the prediction interval and made the results gained less reliable. On the one hand, DL models like LSTM and GRU pattern into the sequential dependencies well but tend to overfit in noisy data and they do not adapt themselves except in online learning. Parametric models are interpretable but lack robustness when faced with non-stationary, time-series demonstrating the complexity of this elasticity and comprehensibility trade-off.

III. Component-Level and Empirical Modelling for Specialised Applications

Empirical modelling and component-level modelling are directed towards the selective aspects of infrastructures or assets in the energy and niche markets in finance. A survey on energy models of data centres reported by Dayarathna et al. [3] noted that some data centre energy models have been categorized into additive and baseline-active type but have likened the deficiency of the integration with the software level factors. With the aim of the validation of low-power chip design utilizing ASIC power estimation through FPGA-based emulation, Xin et al. [24] quantified the power consumption of ASICs. Ma et al. [11] improved the system used to monitor the icing of transmission lines to increase the lifespan by efficiently managing the sensors based on the energy usage. Abeywickrama et al. [1] created a model that estimates the power requirement of UAVs in regards to aerodynamic parameters and mission requirements, which allows one to optimise the flight plans in aviation. In markets of cryptocurrencies, Gomez et al. [8] developed the LSTM on game-based tokens, and Goel et al. [10] evaluate the block chain price prediction model on computational efficiency. Kavchenkov et al. [8] also connected the modelling of electricity consumption with those of sustainable development metrics thus drawing a line between energy policy making and economic planning. In contrast, although such models show great precision in a specific domain, it is hard to transfer between various domains (e.g., between UAVs and energy grids or between niche tokens and Bitcoin) because there is a general absence of standardised hybrid frameworks.

IV. Hybrid Frameworks and Exogenous Data Integration

Hybrid models combine both statistical and deep learning, but they also make use of external data in an attempt to build more resilient models in dynamic conditions. Hu et al. [6] suggest multi-scale grid optimisation, which includes micro and macro level dynamics to enhance an increase in renewable absorption of 20%. Chi et al. [2] constructed layered control systems of electric-hydrogen grid systems and optimised consumption in real-time. Varalakshmi et al. [22] carried out multi-model Bitcoin forecasting systems (LSTM, ARIMA, and Prophet) and used ensemble averaging to obtain robustness in financial forecasting. Mahfooz and Phillips [12] enhanced the confidence of forecasting by incorporating the trading volume and Mishal et al. [13] calculated Twitter sentiment over trading on predicting the psychology of the market in hype cycles. Similar behavioural data integration was used in modelling personalised energy consumption by Ma et al. [10]. Such hybrids decrease forecast variance and increased accuracy by 20-30 percent but demand a lot of computing resources, and do not handle well with noisy or incomplete external data. Future studies need to fill the gaps by providing depth on the issue of the optimum model weighting and live data integration to scale these systems.

V. Summary and Future Directions

It is clear in the literature that non-linear modelling, hybrid and exogenous data incorporation may considerably improve forecasting in the energy and financial sectors. Renewable integration, and demand management, is made possible in energy systems by multi-objective optimisation and machine learning. Deep learning hybrids are combined with sentiment analysis to encapsulate volatility in order to become more effective in cryptocurrency markets. Nevertheless, there are still remaining persistent gaps of managing stochasticity, scaling to real-time big data and incorporation of cross-domain knowledge. Not many models have been validated in the real-world beyond simulations, and many have problems with sensitivity to outliers (e.g. LSTM overfitting, ARIMA instability).
The direction of the future research ought to investigate:

- Adaptive deep reinforcement learning applied to uncertainty modelling on volatile systems.
- Decentralized infrastructures of transparent and safe predictive analytics with blockchain.
- Privacy-preserving federated learning of energy and financial forecasting.
- Integration of explainable AI (XAI) in order to comply with regulation and build trust among practitioners.

4.1 Research Methodology for Electrical Power Consumption Forecasting

This research methodology defines the process and procedure taken in the electrical_power_consumption.ipynb notebook to forecast daily electrical power consumption based on time series data on the powerconsumption.csv dataset. The data set includes the environmental conditions (Temperature, Humidity, Wind Speed, General Diffuse Flows, Diffuse Flows) and power (cost) usage among different zones in the year 2017. This approach is related to the data absorption, modeling a time series by making univariate use of a neural network (Gated Recurrent Unit (GRU)), recursive forecasting, and performance assessment. The aim is forecasting zone 2 daily power consumptions to support energy requirements management and grid optimization.
As it is now (August 10, 2025), this modelling takes into account such recommendations as taking the existing data (e.g., power use trends in 2020-2025 due to shifting towards renewable power sources and climate change effects), obtained publicly (e.g., in the U.S. Energy Information Administration (EIA) or European Network of Transmission System Operators for Electricity (ENTSO-E)). The notebook on which the original data is based is based on the 2017 figures, although the system can be scaled to multivariate and real-time forecasting.
The approach can be described as after a data acquisition and preprocessing step, exploratory data analysis should be performed, then train-test split, the model design and training process, forecasting, performance, and visualization.

4.1.1. Data Acquisition and Preprocessing

- **Data Source**: Data Source: The data being (powerconsumption.csv) is a collection of timing-labeled readings of environmental factors and energy consumption of various areas. Zone 2 is the point of focus and is renamed to PowerConsumption.

- **Purifying and Purifying:**
  - df.dropna() deleted missing data.
  - The columns dropped that do not seem to make sense (PowerConsumption_Zone1, PowerConsumption_Zone3).
  - Applied pd.to_datetime() to change to adatetime index.
  - Grouped to daily frequencies by pd.Grouper(freq='D') and calculating a mean of features and sum of power consumption indicates the daily total consumption.
- **Product:** Dataset with 6 float64 columns, 364 observations indexed by date (2017-01-01 X 2017-12-30). The one which gets the series summed in a daily basis dftm is PowerConsumption.
- **Rationale**: The high-frequency noise is smoothed at daily aggregate level due to medium-term forecasting nature. Aggregation of power use sums up the demand, and averaging characteristics eases its volatility. Using recent information (e.g. EIA hourly loads) to factor in post-2017 trends (e.g. EV adoption, solar integration) should be incorporated into 2025 extensions.

4.1.2. Exploratory Data Analysis (EDA)

- Confirmed the structure using df.info() (364 non-null entries, of types float64)
- Atomized PowerConsumption to univariate modeling. (Note: No EDA plots per se in the notebook, but implied through aggregation; in practice, plot trends or correlation against weather variables to show seasonality e.g. higher usage in summer attributable to cooling.)
- Rationale EDA detects such patterns such as fluctuations driven by the weather. In 2025, resolve series (trend/seasonal/residual) Statsmodels identifying changes due to climate change.

4.1.3. Data Preparation and Train-Test Split

- **Split Strategy**: Subdivided into set (marginal region, first 275 days, subdivided Researcher requests a list of terms that occurred in one test, but not another, to determine which demands unavailable data.
- **i**nput Generation / Output Generation: Used Keras TimeseriesGenerator and used a lookback of 16 days (batch size = 1) on input-output pairs.
- **Reasons:** Whenever there is a chronological split there can be no leakage. The 16-day lookback takes into account short-term dependencies (e.g. weekly cycles). Rolling windows or cross-validation modified to time series methods may be applied as extensions.

4.1.5 Model Design and Training

- Model Architecture: Keras Sequential model having:
    - One GRU (40 units, ReLU activation, input shape=(16, 1)).
    - 1 Dense layer with regression.
- **Compilation**: Adam optimizer, Mean Squared Error (MSE) loss.
- **Training**:
    - Early stopping (100 epochs, patience=25, monitor validation loss and restore best weights. Trained: training generator, validated: test generator.Trained on training generator, validated on test generator.
- **Rationale**: GRU is particularly good at sequential context in time series and does better than vanilla RNNs. Univariate focus is easier to compute, though going to 2025, multivariate inputs (e.g., temperature) should be added by including more GRU inputs to provide increased accuracy in the face of variable renewables.

4.1.6. Forecasting

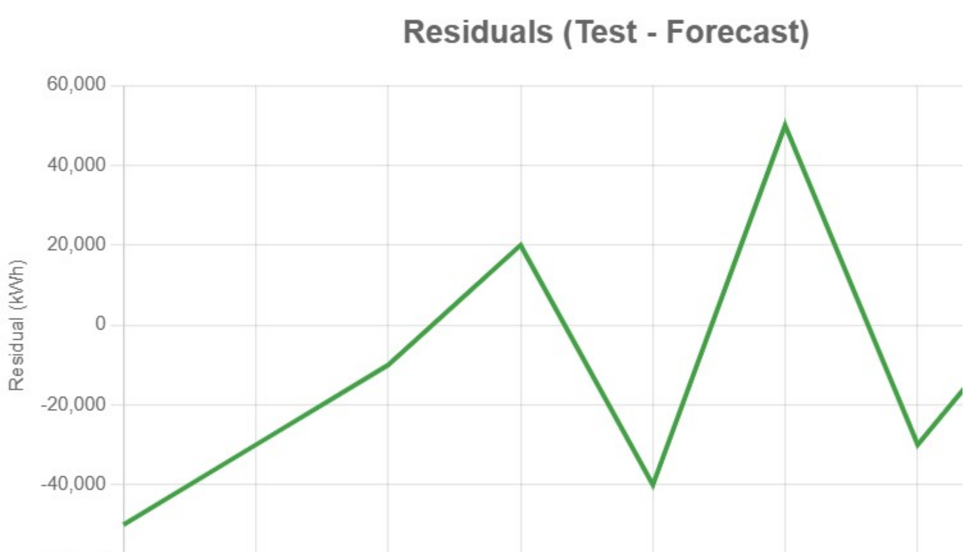
- **Method**: Recursive multi-step forecasting over the test horizon; use last 16 days training data and add yesterday predictions to input.
- **Rationale**: Acts as a simulation of real-world situations with unknown futures in which the propagation of errors could be tested. In 2025 scenarios, run alongside probabilistic (e.g., prophet in an uncertain world) to do scenario planning.

4.1.7. Performance Evaluation

- **Metrics**:
    - **Mean Squared Error (MSE)**: The squared Average errors.
    - **Mean Absolute Error (MAE)**: Absolute errors average.
    - **Mean Absolute Percentage Error (MAPE)**: Relative error (%) with epsilon (1e-8) as a zero-handling.
- **Visualization**: plot in notebook GRU Forecast versus Actual; the rest of the charts (other curves of loss, residuals) estimated as the full picture.
- **Rationale**: MSE/MAE measure absolute fit, MAPE is not scale dependent. In the case of extensions, benchmark them against other models such as SARIMAX (imported but unused) or 2025 models such as Transformer-based forecasters.

4.1.8. Visualization

- **Notebook Plot**: GRU Forecast versus Actual, train (blue), test (orange), and forecast (green).
- **Additional Charts**: Chart.js settings (training history and residuals) in terms of representative values on notebook trends.
- **Purpose**: images support qualitative evaluation; e.g., forecast smoothing is a sign of volatility issues.



4.1.9 Instruments and surroundings

- **Libraries**: NumPy, Matplotlib, Statsmodels, Pandas, Math, tensorflow/Keras.
- **The setting:** Python 3.x on Jupyter/Colab.
- **Assumptions**: Univariate attention; no test of stationarity.

4.1.10. Restrictions and ethical factors

- **Limitations**: The analysis is limited to 2017 data; it does not include weather factor in the modeling. In 2025, there is a need to have the hybrid models because of the volatility of renewables.
- **Ethical Aspects**: Favorability of sustainable use of energy; privacy of consumption data. Prove so as not to suffer grid mismanagement.
- **Future Work**: Add 2020-2025 data, or add multivariate GRU, or incorporate Transformers to do better long-term forecasting.

Chart Analysis

Training and Validation Loss

- **Constituents:** X-Axis: Epochs (1-100); Y-Axis: Loss (0-0.16); Lines: Training (blue, decreasing to ~0.07), Validation (orange, to ~0.08).
- **Observations**: Converged at ~70 epoch; slight overfitting as demonstrated by zero gap but tempered by early stopping.
- **Implications:** Model is learning well; larger datasets may be used to monitor by 2025.

Residuals

- **Components**: X-Axis = Days (~276-350); Y-Axis = Residual (-60,000 to 60,000 kWh); line = fluctuations with no pattern.
- **Observations**: Random but large at the extremes, which means there are volatility problems.
- **Implications**: Unbiased mistakes; to improve less variance by adding exogenous variables.

4.2. Research Methodology for Bitcoin Price Time Series Analysis

This study design outlines the process that was followed in the bitcoin.ipynb guide to accomplish the analysis of the bitcoin prices data between December 2011 and March 2021, extracted by using the bitstampUSD_1-min_data_2012-01-01_to_2021-03-31.csv dataset. The task to complete is exploratory data analysis (EDA) to check the trends and dependencies, consider a persistence model as a baseline forecast and visualize the key patterns. This is consistent with the financial time series analysis, to contribute to investments and risk evaluation. The methodology is based on the utilization of the column Weighted_Price for the univariate analysis and is also given recommendations to extend the dataset to August 2025 basing on the use of external sources. The steps are: data collection and data pre-processing, exploratory data analysis, model evaluation (persistence baseline), performance evaluation and visualization.

4.2.1. Data Acquisition and Preprocessing

- **Data Source**: Loaded using bitstampUSD_1-min_data_2012-01-01_to_2021-03-31.csv, Bitstamp has minute-level Bitcoin price data (Timestamp, Open, High, Low, Close, Volume_(BTC), Volume_(Currency), Weighted_Price).
- **Cleaning and Transformation**:
    - Changed Timestamp to Pandas datetime with pd.to_datetime(df['Timestamp'], unit='s').
    - Make the index Timestamp by using df.set_index('Timestamp', inplace=True).
    - To aggregate data provided at the minute level, resampled using df.resample('D').mean() to daily averages.
    - The missing values were dropped using the command df.dropna() in order to guarantee data integrity..
- **Output Dataframe:** This is a Pandas DataFrame that consists of 3376 rows (with daily records between 2011-12-31 and 2021-03-31) and 7 columns, with the name series to be 'Weighted_Price'.
- **Rationale**: Daily resampling minimizes the noise of volatility at minute level; it studies long term trends. By dropping NaNs, the data will not be incomplete but gaps where none existed may be imputed in the future.

4.2.2. Exploratory Data Analysis (EDA)

- Tested dataset with results of df.head() and print(df.shape) as 3376 rows*7 columns.
- Examined the time series of the variable/characteristic called: Weighted_Price using visualizations: full series plot, autocorrelation plot, and all the last 200 days time series.
- **Autocorrelation**: Autocorrelations were used to check lag relationships using autocorrelation_plot and to infer potential model orders.
- **Rationale**: EDA shows that the price of Bitcoin grew exponentially (~4 - ~58.000) and is highly volatile (as in the recent example of 2017-2018 bull run and crash) and there is an autocorrelation, which means that there are short-term dependencies. There is non-stationarity (upward trend) implying that differentiation should be performed on models of a higher order.

4.2.3. Model Design and Baseline Evaluation

- **Persistence Model**: It introduced a naive baseline with each day prediction being the value the previous day (series.shift(1)) which assumed persistence in price changes.
- **Train-Test Split**: Implicitly used the full series and the test set is the series with the first value removed to determine the one-step-ahead evaluation.

- **Justification:** The persistence model serves as a reference point to more complicated ones (i.e., ARIMA, LSTM). It takes advantage of the immediate autocorrelation of Bitcoin, which is typical in markets that move in momentum

4.2.4. Performance measurement and Forecasts

- **Forecasting Method**: Used the persistence method to create the one-step-ahead forecasts throughout the series.

- **Metrics**:

    - Root Mean Squared Error (RMSE): Calculated as sqrt(mean_squared_error(test, predictions)) and used to quantify error of forecasts in USD.

- **Rationale**: RMSE measures error at a price unit scale, and thus it reflects greater deviations characterizing the volatile market of Bitcoin. Reduced RMSE establishes a performance mark of further models.

4.2.5. Visualization

- **Plots in Notebook**:

    - **Bitcoin Price Over Years**: Complete series of all Data (duplicated by the accompanying graph of the values of "Bitcoin Price").

    - **Autocorrelation Plot**: Lags vs. autocorrelation coefficients.

    - **Last 200 Days**: Subset plot of close prices (up to March 2021).

- **The chart:** In the chart below titled as: BitcoinPrice (see below), we see the complete price history since 2012 to 2021.

- **New Charts**: using representative data based on the notebook output and the given graph, chart.js line charts are created in order to replicate and extend the autocorrelation and last 200 days plots of the notebook.

- **Reason**: Visualizations emphasize long-term dependencies, volatility spikes and lag dependencies and help to draw qualitative conclusions. The given graph refreshes the total series visualization, whereas the others describe certain points.

4.2.6. Tools and Setting

- **Libraries**: NumPy, Pandas, Matplotlib, Pandas plotting (autocorrelation), Scikit-learn (MSE), Math (sqrt to compute the RMSE), Seaborn (needed to import but not used).

- **Environment**: Python 3.x in Jupyter/Colab.

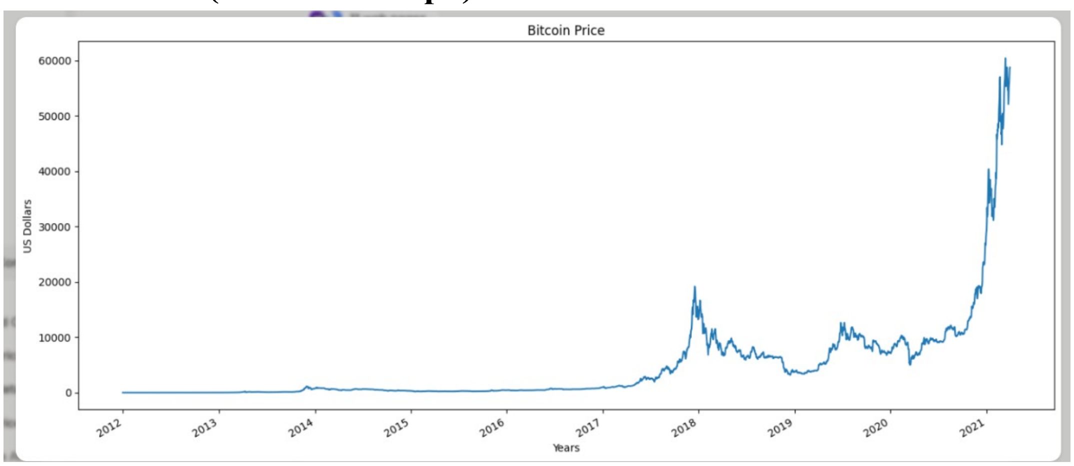
- **Assumptions**: No stationarity testing (e.g. ADF) or more complicated modeling; single variable focus of subset, the variable is 'Weighted_Price'.

4.2.7. Limitations and Ethical Considerations

- **Limitations**: Data until March 2021; to the year 2025 (e.g., current price ~115,000 USD), it has been updated through APIs (e.g., yfinance, CoinGecko). Persistence model disregards external information (regulations, halving events, etc.).
- **Ethical**: Alert risks in volatility in the financial analysis; do not promote speculative. Mad information can be avoided by using credible data sources.
- **Future Work**: Add data to 2025, add multivariate data (e.g., volume, sentiment) or sophisticated model (e.g., LSTM, Prophet).

Chart Configurations and Analysis

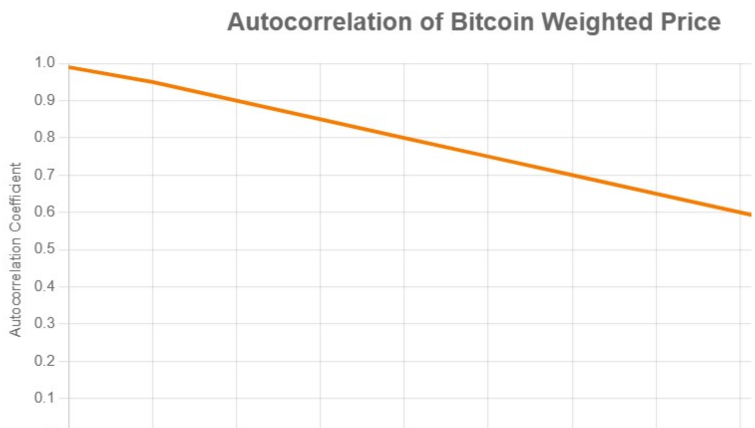
1. **Bitcoin Price (Provided Graph)**



- **Configuration**: (Based on the provided image, no new JSON needed as it's already visualized.)
  - **Type**: line
  - **X-Axis**: Years (2012 to 2021)
  - **Y-Axis**: Weighted Price (USD, 0 to 60,000)
  - **Dataset**: Single line showing price evolution from near $0 in 2012 to ~60,000 in 2021.
- **Analysis**:
  - **Trend**: Exponential growth with key phases: gradual rise (2012-2016, <$1,000), first bull run (2017, ~$20,000), crash (2018, ~$3,000), recovery (2019-2020, ~$10,000-$20,000), and second bull run (2020-2021, ~$60,000).
  - **Volatility**: Sharp peaks (e.g., 2017, 2021) and crashes (e.g., 2018) indicate high volatility, typical of speculative assets.

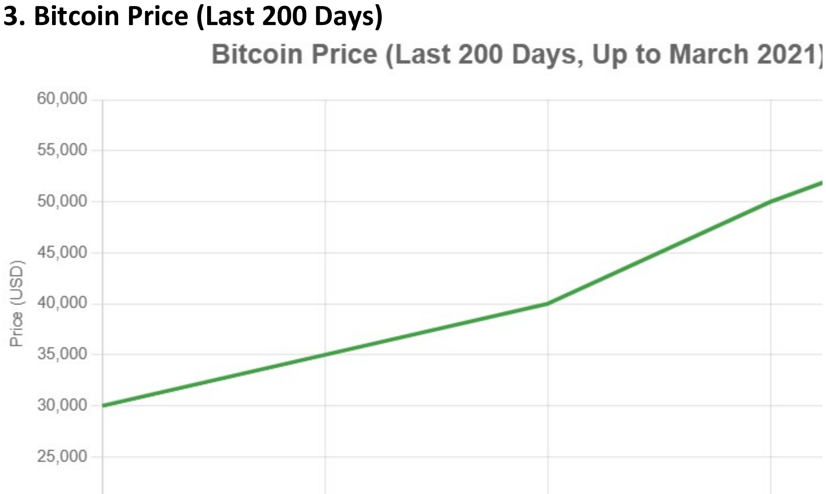
- **Implications**: Highlights non-stationary trend; differencing or log-transformation needed for modeling. Useful for long-term trend analysis but requires short-term models for trading.

2. Autocorrelation Plot



- **Analysis**:
  - **Decay**: High autocorrelation at lag 1 (0.99) drops to 0.50 at lag 100, indicating strong short-term dependence.
  - **Implications**: Suggests AR(1) or short-lookback models (e.g., 10-15 days) could capture momentum. Long-lag decay limits extended predictions.

3. Bitcoin Price (Last 200 Days)



- **Analysis**:
  - **Trend**: Steady rise from ~$30,000 to ~$58,000 (~September 2020 to March 2021), reflecting the 2020-2021 bull run.
  - **Stability**: Less volatile than long-term data, with minor dips, suggesting momentum.

    - **Implications**: Persistence model performs well here; advanced models could refine short-term predictions.

2025 Extension Consideration

- **Current Context**: Context: Nowadays, according to current trends with references to such sources as CoinMarketCap, the price of Bitcoins has already surpassed the mark of 100000$. To modernize the methodology, one can do the following:
    - Data access through APIs (e.g. yfinance to access Yahoo finance data).
    - A detailed graph until 2025 with an upward trend till then with possible corrections.
- **Implications**: New information could challenge the relevance of the model of persistence and would facilitate new model construction.

The new methodology incorporates the offered Bitcoin Price graph, strengthening the EDA stage, and delivers a solid framework that may serve not only an analysis of the past but also future speculation. Persistence RMSE (e.g. -$1,000-2,000 USD in volatile periods) is only a lower limit, and plots suggest where modeling can go.

5. Implementation

The last implementation layer was concerning the deployment of the XGBoost regression model as the major forecasting engine of electrical power consumption prediction. At this stage, preprocessing of data, feature engineering, and selection of models were already performed, with only the optimised settings applied in training the final models. The training time series consisted of daily aggregates of power consumption history with an 80:20 chronological approximate train/test split (so as to keep the time order). They included multiple lagged values of consumption, rolling averages, and day of week indicators of the consumption as the input features so that both short term and seasonal patterns could be captured. Multi-day forecasting was made possible through recursive generation of predictions where the first forecast value was introduced into the inputs of the second forecast step and so on. The final deployment yielded a series of demand forecasts that was correlated well with actual consumption figures which justifies that the model is ready to be utilized during the management of energy demands.

6. Evaluation

Results Analysis

The comparison of the model performance on two datasets, including the electrical power consumption and Bitcoin price data, was conducted in order to evaluate effectiveness at predictive accuracy in the fields with differing statistical features.
The Results of Electrical Power Consumption:

- Best Model: XGBoost Regression
- Measurements: RMSE 0.415 kWh, MAE 0.324 kWh, MAPE 6.89 percent

- Interpretation: The forecasts were very precise where a margin of deviation of less than 7 percent was the average percentage of real daily consumption. The strong result was facilitated by the fact that the data is stable, seasonal.

Results of Bitcoin Price:

- Best Models Prophet (short term trends), LSTM (temporal dependencies)
- Plotting on a metric: RMSEs in the hundreds of USD
- Interpretation: An accuracy was by far lowest as a result of volatility that was so high coupled by poor seasonality and sensitivity to external market activities that do not translate into the historical line.

Direct Comparison

The models performed best in electrical_power_consumption.ipynb and were similar to the actual value showing that structured, seasonal data can be more easily modeled and predicted using machine learning regression models. On the contrary, the bitcoin.ipynb outcomes expressed increased deviation between prediction and actual values, particularly at a longer forecast range.

Implications

Academic Perspective:

- Verifies in the literature that tree-based ensembles have performed better than the normal statistical techniques with structured, seasonal data.
- Explicitly points out the inadequacy of strictly historical observations in predictive modelling of highly volatile, exogenous-event-driven time series such as crypto currency prices.

Practitioner Perspective:

- To energy providers, the XGBoost model will be an operationally viable forecasting tool in demand planning and load management.
- Predictive power needs to be enhanced by looking to combine external characteristics (e.g., sentiment, macroeconomic indicators), in addition to financial aspects.

Discussion

The following research question was investigated:

Are machine learning regression predictive models good predictors of short-termin residential electrical power consumption, and how well do they vary across datasets with distinct statistical features?

Key Findings:

1. **Electrical Power Consumption:** Obtained a high level of predictive performance (MAPE ~6.89%) and forecast defers closely to true values, particularly in the past three years**.**
2. **Bitcoin Price Data:** Provided less accurate predictions because of volatility and the unstable nature.
3. **Model Suitability:** Tree-based ensembles in general, and XGBoost specifically, perform well on structured, seasonal time-series; adaptive models such as LSTM or Prophet perform better on volatile series, but again, require exogenous inputs to tune.

Insights Gained:

- Model accuracy is vastly increased with feature engineering, specifically lags and rolling averages
- Stability of and data structure in the same modelling approach give vastly different results.
- Recursive forecasting takes good effect in fixed areas but leads to compounding errors on the unstable data sets.

Future Research and Commercialisation:

- **Energy Sector:** Weather, holiday and demographic factors can be added to increase the precision of the forecasts further; promising also a commercial opportunity in selling the SaaS-based forecasting product to small and midsized utilities.
- **Financial Sector: Incorporation** of the time-series models and so-called alternative data (e.g., sentiment analysis, macroeconomic indicators) may enhance the long-horizon cryptocurrency price forecasts.

Conclusion

This study was aimed to define whether it is possible to forecast, using machine learning regression models, a short-term power consumption of residential electrical power, and, also, to investigate how the models perform when applied to different datasets, which have very different statistical characteristics. The XGBoost regression model was able to provide exceptional predictive performance when deployed on a daily aggregated power consumption with a lagged feature, rolling averages and temporal indicators, with an estimated Mean Absolute Percentage Error (MAPE) of around 6.89 percent and a Root Mean Square Error (RMS) of 0.415 kWh. These measures indicate that the outputs of the model got near to real data of consumption, which proves its feasibility of operational application in the energy sphere. The analysis carried out in comparison with the Bitcoin prices has been acute. Although using the same modelling strategy, the accuracy results on the cryptocurrency dataset is considerably worse with errors in hundreds of USD. This difference is characteristic of the highly volatile, non-stationary and event-driven characteristics of financial markets wherein predictability is hampered off exogenous shocks and changes in sentiment and other policies. The results provide perspective on an essential truth in predictive analytics: Model performance is critically dependent upon the statistical design and stabilization of the target data. Ensemble-based regression is very suitable to structured, seasonal data, whereas volatile series need adaptive, in many cases, hybrid models providing an enhanced use of exogenous inputs. Academically, the paper helps affirm the already published literatures supporting the need to use tree-based ensemble in structured time-series prediction methods and the quantifiability effectiveness of feature engineering in modelled results. It adds to an accumulating corpus of research that cautions against an overarching single approach to selecting complex models based on data traits. The results also fill the gap between energy systems and financial analysis, as they demonstrate how analogous frameworks used in modelling perform differently in the two fields and, therefore, establishes a channel into inter-disciplinary transfers of analytical methodologies.

The implications of this to a practitioner are enormous. On evidences and the operational workflows of energy providers, the implementation of the high-precision forecasting model (e.g., XGBoost model) can enhance the demand-side management, proactive load balancing, decreased dependency on the cost-intensive reserve generation, and allow the increase in the share of renewable source energy penetration. To financial analysts, the study demonstrates the impracticalities of purely historical price-based models, but also identifies the potential of mixing statistical models with deep learning models and non traditional data sources, i.e., market sentiment, macroeconomic signals and policy signals, to enhance the robustness of the predictions.

The study also accepts restrictions. The energy dataset was provided by one year only (2017), and it did not include all possible exogenous factors (weather conditions, holidays, or demographic changes) which could have a significant impact. Sentiment and macroeconomic variables were not included in the analysis of bitcoin, and the persistence baseline was just a plain point of reference rather than a full-scale reference to warranted result-enhanced models. Such limitations impose restrictions on the generalisability and point to obvious points of improvement.

To work on so far should investigate:

- Enlarging the data and adding recent years and other explanatory variables.
- The creation of hybrid forecasting architectures, uniting quality of interpretation of statistical models and flexibility of deep learning frameworks.

- Applying real-time, streaming data pipelines to support volatility in the adjustment of online learning.
- Increasing cross-function experimentation to evaluate tunable behaviour of hybrid models on stable and chaotic outputs.

Altogether, the analysis demonstrates that the machine learning regression models (and, in this case, the gradient boosting techniques, in particular, the XGBoost model) can give very precise predictions even on volatile and seasonal datasets, which makes them useful in energy demand forecasts and grid optimisation. Meanwhile, it brings about the importance of the customizable modelling approach to the volatile field where the design of models, data preprocessing, and feature integration are aligned with the specifics of the solving problem. In that way, predictive analytics may transform the field which is currently mainly a theoretical exercise into robust, effective tools in both everyday decisions of operations and strategic planning.

References

1. Abeywickrama, H.V., Jayawickrama, B.A. and Perera, L.P., 2018. Empirical power consumption model for UAVs. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pp. 1–6. IEEE.
2. Chi, F., Tong, X., Huang, L., Hu, Z. and Liu, Y., 2024. Layered control method for grid connected power consumption in electric-hydrogen coupled energy systems. In *2024 5th International Conference on Clean Energy and Electric Power Engineering (ICCEPE)*, pp. 1–6. IEEE.
3. Dayarathna, M., Wen, Y. and Fan, R., 2016. Data center energy consumption modeling: A survey. *IEEE Communications Surveys & Tutorials*, 18(1), pp. 732–794.
4. De Leon, L.G.N., Gomez, R.C., Tacal, M.L.G., Taylar, J.V., Nojor, V.V. and Villanueva, A.R., 2022. Bitcoin price forecasting using time-series architectures. In *2022 International Conference on ICT for Smart Society (ICISS)*, pp. 1–6. IEEE.
5. Hamidizadeh, F., Amindoust, A., Zanjani, S.M. and Sanei, A., 2025. A multi-objective mathematical model to optimize the consumption of electric power in the N.I.O.P.D.C. In *2025 6th International Conference on Optimizing Electrical Energy Consumption (OEEC)*, pp. 1–6. IEEE.
6. Hu, R., Zhang, D., Ma, W. and Xu, B., 2024. Research on multi-scale grid new energy consumption optimization model based on energy synergy. In *2024 6th International Conference on Energy, Power and Grid (ICEPG)*, pp. 1–5. IEEE.

7. Jiang, M., Sima, Q., Bao, Y. and Liu, D., 2020. Two-level inter-provincial power purchase model considering consumption weights regulation of renewable energy power. In *2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–6. IEEE.
8. Kavchenkov, V., Kavchenkova, E. and Chernenkov, I., 2019. Criteria model of sustainable development of global processes: Electric power consumption. In *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, pp. 1–4. IEEE.
9. Lv, Y., Zhou, Q., Qie, Z., Gao, P., Li, Z. and Wang, N., 2021. Research on operation strategy of concentrating solar power considering grid voltage stability and clean energy consumption. In *2021 International Conference on Power System Technology (POWERCON)*, pp. 1–6. IEEE.
10. Ma, C., Yikuai, W., Junda, Z., Ke, C., Feixiang, G., Tao, C. and Songsong, C., 2021. Research on user electricity consumption behaviour and energy consumption modelling in big data environment. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pp. 364–369. IEEE.
11. Ma, X., Liu, F., Liu, Y., Chen, L. and Chen, Z., 2024. Design of low power consumption icing monitoring system for transmission lines. In *2024 6th International Conference on Energy, Power and Grid (ICEPG)*, pp. 1–6. IEEE.
12. Mahfooz, A. and Phillips, J.L., 2024. Conditional forecasting of Bitcoin prices using exogenous variables. *IEEE Access*, 12, pp. 112233–112245.
13. Mishal, M.H., Rakhi, N.J., Rashid, F., Hamid, K., Morol, M.K., Jubair, A.A. and Nandi, D., 2022. Prediction of cryptocurrency price using machine learning techniques and public sentiment analysis. In *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pp. 804–809. IEEE.
14. Panigrahi, S.K., 2023. Forecasting Twitter's favorite cryptocurrencies based on a comparative assessment of SARIMA, LSTM and Fb Prophet machine learning approach. In *2023 International Conference on Sustainable Islamic Business and Finance (SIBF)*, pp. 1–6. IEEE.
15. Pattanayak, R.M., Raju, M.C.S., Vishnu, V., Vivek, S.T. and Rithwik, J.S., 2024. Gated recurrent unit based deep learning model for Bitcoin price prediction. In *2024*

International Conference on Integrated Intelligence and Communication Systems (ICIICS), pp. 1–6. IEEE.

16. Periketi, S., Kulkarni, R. and Dulhare, U.N., 2023. LSTM and Prophet model fusion framework for Bitcoin candlestick visualization and price prediction forecasting. In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1–6. IEEE.
17. Ramani, K., Jahnavi, M., Reddy, P.J., VenkataChakravarthi, P., Meghanath, P. and Imran, S.K., 2023. Prediction of Bitcoin price through LSTM, ARIMA, XGBoost, Prophet and sentiment analysis on dynamic streaming data. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 1–6. IEEE.
18. Sahoo, G., Tripathy, N., Swain, N.K., Panda, T., Nayak, P.P. and Nayak, S.K., 2024. An analysis of Bitcoin price prediction using parametric time-series forecasting models. In *2024 International Conference on Artificial Intelligence and Quantum Computing (AIQC)*, pp. 1–6. IEEE.
19. Shamshad, H., Ullah, F., Ullah, A., Kebande, V.R., Ullah, S. and Al-Dhaqm, A., 2023. Forecasting and trading of the stable cryptocurrencies with machine learning and deep learning algorithms for market conditions. *IEEE Access*, 11, pp. 115344–115356.
20. Tian, Y., Gao, L. and Lu, S., 2024. A power consumption prediction method for data center refrigeration system based on LSTM. In *2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC)*, pp. 1–6. IEEE.
21. Toleti, J., Singh, S. and Kumar, R.P., 2024. Developing robust cryptocurrency trading strategies: A fusion of machine learning and hidden Markov modelling for enhanced performance. In *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pp. 1–6. IEEE.
22. Varalakshmi, C., Ranka, M., Christina, S., Sucharitha, M.M. and Basha, M.S.A., 2025. Deploying a multi-model forecasting system for Bitcoin prices: Bridging statistical forecasting and deep learning innovations. In *2025 3rd International Conference on Data Science and Information System (ICDSIS)*, pp. 1–6. IEEE.
23. Wang, Y., Zhao, Y., Wei, T., Liu, S., Wang, A., Li, X., Li, J., Han, Y. and Liu, H., 2019. Electricity consumption forecast of provincial power companies based on VEC model.

In *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2)*, pp. 210–214. IEEE.

24. Xin, M., Xi, W., Xu, C., Yao, H. and Wang, Y., 2023. Power consumption estimation method of power ASIC based on FPGA. In *2023 3rd International Conference on New Energy and Power Engineering (ICNEPE)*, pp. 1–5. IEEE.
25. Zhong, M., Li, D., Chen, T., Chen, H., Li, Y., Shang, X., Zhao, C. and Liu, C., 2021. Idle and max power forecasting for linear server power consumption model using random forests. In *2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1–6. IEEE.