UGANDA CHRISTIAN UNIVERSITY

A Center of Excellence in the Heart of Africa

DEPARTMENT OF COMPUTING & TECHNOLOGY

Faculty of Engineering, Design & Technology

Time Series Analysis and Forecasting

Report

Presented by:

Name	RegNo	AccessNo
Lorraine Paula Arinaitwe	M23B38/004	B20729

April 23, 2025

Abstract

This study investigates the application of time series forecasting techniques on two distinct and rich datasets: **household power consumption** and **stock prices**. The goal is to build accurate and interpretable predictive models through thorough preprocessing, exploratory analysis, and the implementation of statistical, machine learning-based, and deep learning forecasting methods.

The household energy consumption dataset, sourced from Google Drive, contains minute-level electrical usage data recorded over nearly five years. Preprocessing steps included converting the raw text data into a structured format using Pandas, handling invalid entries such as `?`, and imputing missing values through time-based interpolation (forward and backward fill). The `Date` and `Time` columns were combined to create a unified `datetime` index, which was then used to resample the data from minute-level to daily frequency—reducing noise and better reflecting long-term consumption trends. Exploratory data analysis identified `Global active power` as the most suitable target variable, supported by strong correlations with `Global intensity` and the sub-metering features. Outliers were detected and treated using the Interquartile Range (IQR) method. The series was evaluated for stationarity using both the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. Seasonal decomposition and autocorrelation analysis (ACF and PACF) were conducted to uncover temporal patterns. Finally, time-based features such as lag values and rolling averages were engineered, and the dataset was split into training (80%) and testing (20%) subsets for model development.

The second dataset consists of S&P 500 stock prices sourced from Kaggle, containing daily stock data for 478 companies. The data was initially cleaned by handling missing values through forward fill and setting a multi-index using 'date' and 'Name' (company ticker). For modeling purposes, two representative companies were selected for in-depth analysis. Seasonal decomposition and multivariate correlation analysis via heatmaps were used to uncover relationships between features. ADF and KPSS tests revealed non-stationarity, which was corrected using first-order differencing. Lagged values and rolling statistical along with Date-based features were engineered for improved predictive performance. The processed data was split into training and test sets using an 80/20 ratio.

These rigorous pre-processing frameworks laid the groundwork for deploying statistical forecasting models such as Moving Average, Exponential Smoothing, and ARIMA, as well as advanced models including Random Forest and Long Short-Term Memory (LSTM) neural networks. Model performance was evaluated using metrics such as Root Mean Square Error(RMSE), Mean Absolute Error(MAE), and Mean Absolute Percentage Error(MAPE). Among all methods, the Random Forest model yielded the highest accuracy for power consumption forecasting, while Exponential Smoothing performed best for stock price prediction. The findings demonstrate the value of comprehensive pre-processing and hybrid modeling approaches in generating reliable forecasts across multiple domains.

Introduction

The selection of best-performing models was based on their ability to minimize forecasting error on the test datasets while maintaining generalizability. The results below highlight the top-performing models for each task, followed by a comparative analysis of alternative approaches.

1. Best-Performing Models & Evaluation Metrics

Household Power Consumption

Random Forest model provided the best results:

- **MAE**: 0.0107
- **RMSE**: 0.0086
- **MAPE**: 0.96%
- Generalized well and outperformed traditional models.

Other models:

- **Exponential Smoothing (Holt-Winters) Trend: mul, Seasonal: add, Period: 21**: RMSE 0.1908, MAE 0.1424, MAPE 16.51%
- **Moving Average (Window=28)**: RMSE 0.1374, MAE 0.1048, MAPE 10.48%
- **ARIMA (1, 0, 1) (AIC=-248.76)**: RMSE 0.1743, MAE 0.1214, MAPE 14.24%
- **LSTM**: RMSE 0.1942, MAE 0.1500, MAPE 16.87%

Stock Prices (Zion Company)

ARIMA (1,0,0) (AIC=3169.38) was most effective:

- **RMSE**: 1.9551
- **MAE**: 1.8575
- **MAPE**: 4.18%

Other models:

- **Exponential Smoothing (Holt-Winters)**: RMSE 1.4629, MAE 1.2033, MAPE 2.76%
- **Moving Average (Window=35)**: RMSE 1.9193, MAE 1.8243, MAPE 4.05%
- **Random Forest**: RMSE 3.9595, MAE 2.7355, MAPE 5.65%
- **LSTM**: RMSE 9.6349, MAE 6.0114, MAPE 12.29%

2. Forecast Insights & Implications

(a). Power Consumption

- **Peak Usage**: Evenings (6–9 PM), aligned with household activities.
- **Seasonality**: Higher energy use during colder months.
- **Weekly Cycles**: Clear recurring usage patterns.

Implications:

- **Grid Optimization:** Accurate demand forecasts enable utility providers to optimize energy distribution and reduce strain during peak hours.
- **Dynamic Pricing:** Enables the implementation of time-of-use pricing schemes to shift consumer behavior and reduce peak loads.
- **Energy Conservation:** Forecasting tools empower consumers to monitor and adjust their usage habits, contributing to sustainability goals.
- **Policy Planning:** Government agencies can use consumption forecasts to design targeted energy-saving policies and subsidies.
- **Renewable Integration:** Aligning demand forecasts with renewable energy availability (e.g., solar/wind) supports cleaner grid management.

(b). Stock Prices

- Strong persistence with **slow ACF decay**.
- AR(1) pattern post-differencing.
- Identified seasonality and volatility patterns.

Implications:

- **Investment Strategy:** Improved forecasting aids in building risk-adjusted strategies and maximizing return on investment.
- **Portfolio Management:** Time series models can help balance portfolios by predicting underperforming or overperforming stocks.
- **Automated Trading:** Forecasting algorithms can be integrated into algorithmic trading systems for real-time decision-making.
- **Market Risk Mitigation:** Early detection of trend shifts or volatility can support better hedging strategies and reduce losses.
- **Macroeconomic Forecasting:** Stock predictions can serve as proxies for broader economic indicators and inform corporate strategy.

3. Challenges Faced

- **Missing/Invalid Values**: Required custom parsing and interpolation.
- **High Granularity**: Needed resampling from minute to daily frequency (for Power consumption). Multi-indexing for stock prices dataset.
- **Non-stationarity**: Addressed through differencing and transformations.
- **Model Complexity**: LSTM needed significant tuning.
- **Computation Time**: Forecasting models were resource intensive.

4. Future Recommendations

- **Incorporate External Features**: Weather, macroeconomic variables.
- **Model Automation**: Leverage grid/random search and AutoML.
- **Real-Time Forecasting**: Implement streaming data pipelines.
- **Robustness**: Use ensemble methods, fallback models, and monitoring.
- **Deployment**: Flask dashboards and periodic model retraining.