Identifying Unusual Human Movements Using Multi-Agent and Time-Series Outlier Detection Techniques

Conference Paper · February 2023

DOI: 10.1109/ICARC57651.2023.10145617

3 authors, including:

PPG Dinesh Asanka
University of Kelaniya

Chathura Rajapakse
University of Kelaniya

Identifying Unusual Human Movements Using Multi-Agent and Time-Series Outlier Detection Techniques

PPG Dinesh Asanka
Department of Industrial Management
University of Kelaniya
Kelaniya, Sri Lanka
dasanka@kln.ac.lk

Chathura Rajapakshe
Department of Industrial Management
University of Kelaniya
Kelaniya, Sri Lanka
chathura@kln.ac.lk

Masakazu Takahashi
Graduate School of Innovation and Technology Management,
Yamaguchi University
Yamaguchi, Japan
masakazu@yamaguchi-u.ac.jp

2023 3rd International Conference on Advanced Research in Computing (ICARC) | 979-8-3503-4737-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICARC57651.2023.10145617

***Abstract*— This research paper has introduced knowledge-driven multi-agent technology for automated machine learning in time series analysis in the context of human mobility. The main objective of this research is to identify unusual human mobility using Time Series outlier detection techniques with a more efficient multi-agent system. Detection of unusual human movement can be helpful for many domains, such as security, marketing, and health. A mobile dataset in Hiroshima, Japan between 2019-December to 2020-November was used for this research. The mobile dataset was converted to time series for multiple locations in Hiroshima, Japan. Since many different parameters are selected for time series, the message space multi-agent technique is used. Sub agents are introduced for duplicate removal, missing data replacement, and outlier detection. Multiple processing agents and a control agent were introduced to predict the missing values to improve the efficiency of the model. Finally, using the Seasonal-Trend decomposition techniques, unusual movements are identified, and unusual human movements are plotted with the holidays. Multiple outlier points were detected for all the locations, and there were more than a hundred outlier points were detected for the selected locations.**

Keywords*— *Time Series Analysis, Human Movements, Outlier Detection, Multi-Agent Technique, Message Space Agent.

I. INTRODUCTION

Unusual human movement identification is important as it is used in many domains. Unusual human detection would be beneficial mainly for security reasons, where authorities can take proactive actions. Furthermore, for a marketing campaign identifying the pattern of unusual human movement is helpful to enhance the effectiveness of the campaign. The health domain is another domain where unusual human movement identification will be benefitted. Identifying unusual human movements will help detect events such as a pandemic. Considering the said requirements, it is important to detect unusual human movements.

Due to the availability of many sophisticated devices, human movement data is captured in various technologies and different media. IoT devices and mobile devices have the capability of capturing human movement. Due to many data points, more sophisticated techniques must be utilized for time series forecasting. Apart from different types of forecasting, there need to be preprocessing tasks for time forecasting such as data preparation, data cleaning, stationary testing, and normalization, missing values identification and correction, detection and correction of the outlier values. In machine learning, automated machine learning techniques are used to accomplish the above tasks. However, a novel method using knowledge-driven multi-agent technology is proposed due to the above complexities. Fig. 1. shows the overview of the multi-agent architecture for time series forecasting.

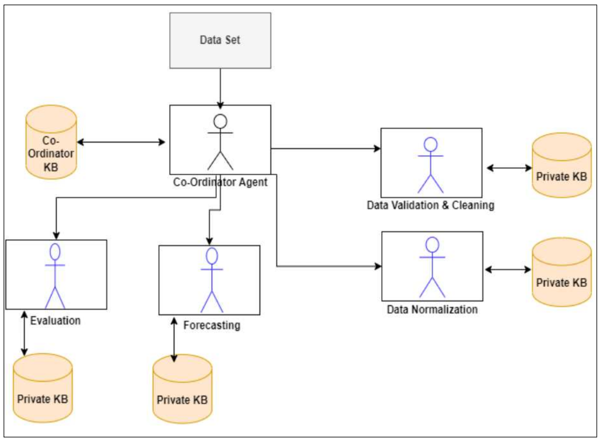

Fig. 1. Overview of Multi-Agent architecture for Time Series forecasting

The main objective of this research is to provide an automated novel machine learning approach to time series forecasting and time series related preprocessing activities such as missing value correction and outlier detection. The proposed multi-agent architecture consists of data validation and cleaning, data normalization, forecasting, and evaluation agents, and these primary agents will have multiple sub-agents. The outlier agent is a sub-agent of the data validation and cleaning, which is the main focus of this research paper.

This research paper exposed a method to identify unusual human movement using a dataset captured using mobile phones, which is part of the more extensive research of time series forecasting using multi-agent, as shown in Fig. 1.

The selected dataset has mobile data collected from people's mobile phones in Hiroshima, Japan, for over one year. This dataset contains data during the covid-19 lockdown period. The main objective of this research paper is to discuss how to convert the existing data into a time series, detect missing values and replace them, and detect the outliers using multi-agent technologies. Since outlier detection requires knowledge-driven decision-making, multi-agent technologies are used. Although this research paper focuses on specific data sets, the proposed multi-agent methodology can be used for many datasets.

This paper is organized into a literature review, which discusses research on time series, outlier detection, and multi-agent implementation. The Methodology section describes the multi-agent model implemented to achieve the said objectives. The Implementation section discusses the implementation and

979-8-3503-4737-1/23/$31.00 ©2023 IEEE

results at different sub-agents, and finally, the Concluding remarks section summarizes the findings and future work.

II. Literature Review

This paper mainly uses multi-agent and time series techniques; hence it was essential to examine the research work on multi-agent and time series. The time series technique was employed in many domains, such as power and energy [1], carbon prices [2], fruit crops production [3], oil production and export [4], weather and climate forecasting [5,6], and many other domains [7,8].

In most cases, the time-series technique was employed to forecast future continuous numerical values such as electricity consumption. This research has used the Augmented Dickey-Fuller technique [9] to determine whether the time series is stationary. In case the time series is non-stationary, the Min-Max technique was employed to convert the time series into stationary [1,2].

Before outlier detection, data cleaning techniques have to be applied. The most concerning operations are stationary testing, identifying the missing data, and replacing missing values. Typically, time series have a high rate of error records due to various reasons [10]. AutoML techniques were introduced by research [11] to clean univariate time series. This research has identified missing data, missing timestamps, outliers, duplicated observations, inconsistent data, problems with data types, and problems with timestamp format as the error data in a time series.

Automated Data Cleaning (AutoDC) [12] technique was introduced to clean the dataset with automated processes. In the proposed AutoDC technique, label correction and Data augmentation are done. After the AutoDC is performed, the improved dataset is presented for forecasting.

Though these techniques have provided some outlier detection levels, they have few limitations in utilizing the best combination. To cover the above research gap, this research has proposed a multi-agent technique. Therefore, multi-agent technologies were examined concerning time series and outlier detections.

Multi-Agent technology is used in many research and there are many implementations for time series forecasting using multi-agent technologies. Preprocessing and aggregation multi-agents were used to extract data from three sources [13]. The multi-agent system was used in another research for forecasting financial time series [14]. Adaptive multi-agents were developed in this research for the selected number of time series techniques with selected parameters. This research has a limited number of techniques and is not extensible.

Azure Machine Learning is a Microsoft-developed cloud platform that supports MLOPs and AutoML [15]. This AutoML platform facilitates forecasting techniques with the selected evaluation techniques. Hence this platform does not provide missing value replacement stationary testing, data processing, and outlier detection. Auto_TS: Auto_TimeSeries [16] is a python library that automatically builds multiple Time Series models. Auto_TimeSeries lets you build and select multiple time series models using techniques such as ARIMA, SARIMAX, VAR, decomposable (trend + seasonality + holidays) models, and ensemble machine learning models.

Furthermore, Auto_TimeSeries technique uses either RMSE or Normalized RMSE as evaluation techniques. However, Auto_TimeSeries expects clean Time series data, which is the research gap that is filled by the proposed multi-agent method in this research. The Azure machine learning (AML) platform has outlier detection capabilities [17], but most configurations should be done manually. AML has used the Martingale framework with six different implementations [18]. Research [19] was done to detect the distributed denial of service attacks using time series anomaly detection techniques. ARIMA, LSTM ensemble technique was proposed in the research. All the research that was discussed have indicated the requirement of a robust and knowledge-driven multi-agent system to detect the outlier in time series.

Research [20] has presented an implementation of an intelligent multi-agent system to troubleshoot issues in a Service (SaaS) Environment using Multi-Agent Technology. This research paper has adapted Message Space Agent (MSA) architecture which is the extension of Blackboard architecture proposed by another researcher [21].

Having identified the research, the main objective of this research is to implement a knowledge driven agent system to detect unusual human mobility using time series techniques.

III. Methodology

The proposed research architect consists of four main components: data processing, missing value detection and replacement, stationary validation, and outlier detection as shown in Fig. 2.

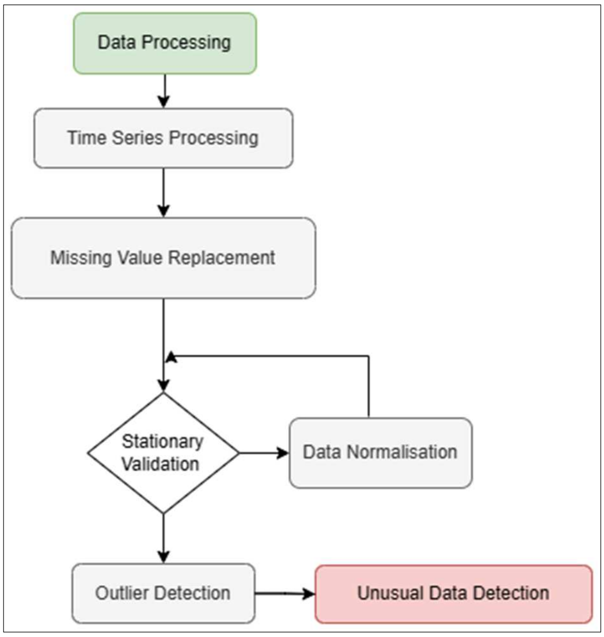

Fig. 2. High level overview of unusual data detection in Time Series

In this method, it is required to convert the state space data to time series that is collected through mobile phones. Due to the large volume of data, removing the duplicate data is required before converting it to time series data. Since this research determines the unusual human movement at the different locations in Hiroshima, Japan, latitude and longitude information should be converted to the time series using geo-functions.

The next step is to identify the missing value and correct them with valid values as time series need continuous data without any missing values. After missing values are detected and replaced, stationary testing should be done before proceeding to the outlier detection.

As indicated, a message space multi-agent mechanism is proposed to detect unusual human movements as shown in Fig. 3. In this multi-agent architecture, data type validation, timestamp format validation, and stationary testing are simple processing agents and do not require any knowledge base. However, all the other agents, missing values, duplicate value detection, data normalization, and outlier detection need their own knowledge bases. These knowledge bases have their own updatable ontology with the knowledge emerging capability.

In the multi-agent architecture in Fig. 3., precedence is set, which means pre-validation, data processing, missing value identification, missing value correction, stationary testing, data normalization, and outlier detection should happen in the order. These tasks cannot run in parallel, and agent architecture should enforce the sequence nature of execution.

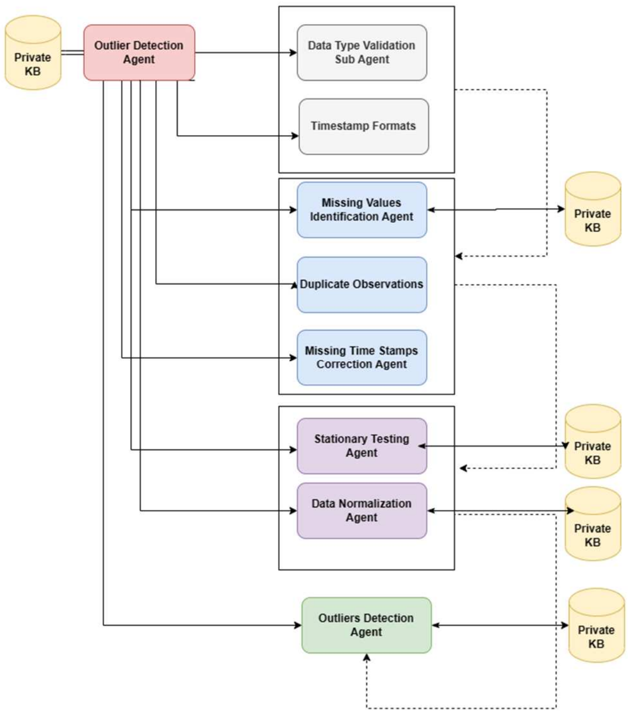

Fig. 3. Proposed Multi-Agent architecture to detect unusual human movements.

The next important component in the proposed multi-agent system is the missing value detection and replacing the missing values with the possible ones. as time series analysis needs a continuous dataset, missing values cannot be ignored and must correct the missing values.

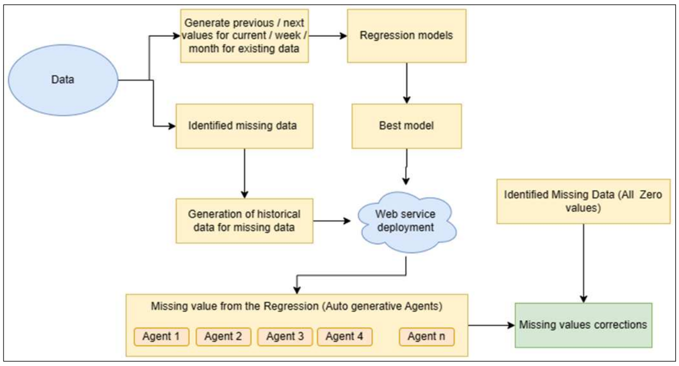

Fig. 4. Detecting missing values and correcting missing values using multi agents

As shown in Fig. 4. regression techniques and the simple rule technique were used to correct the missing values. If all the previous and next values for a given data point are zero, then the missing value will be zero which is corrected by the simple rule technique. The best regression model was selected by comparing the Root Mean Squared Error value. By implementing the regression model, missing values are corrected.

However, due to the high number of missing values in typical time series data and proposed regression techniques, it will take substantial time to correct missing values. Therefore, this research has suggested self-emerging multiple parallel processing agents. Further, these agents are re-startable agents where it can start from where it left off rather than starting from the beginning at every time.

Once missing values are corrected, time-series dataset is available for outlier detection. Since time series are being used for outlier detection, verifying the stationary validation for the derived time series is necessary. Augmented Dicky fuller is proposed for stationary testing.

Out of the available outlier detection techniques, Seasonal-Trend decomposition (STL) using locally estimated scatterplot smoothing (LOESS) is chosen to detect outliers [22], as shown in Fig. 5, STL decomposition.

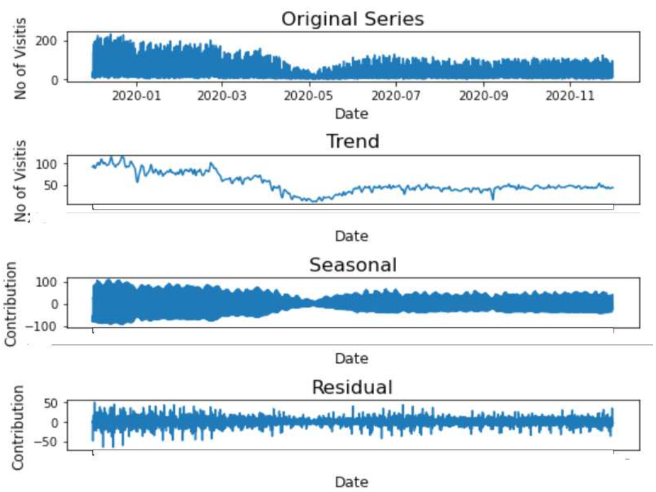

Fig. 5. Decomposition of Time Series using STL (LOESS).

After decomposing time series into trend, seasonal and residual, residual is further analyzed to detect outliers. The range is defined for different k values as shown in EQ 1.

$$\text{Upper range} = \mu + (k * \sigma)$$

EQ (1)

$$\text{Lower range} = \mu - (k * \sigma)$$

Where, μ - Mean for the residual series, σ – standard deviation of the residual, k – different values for band factor that will decide the outliers.

Fig. 6. shows the range when k = 3. By changing the band factor to different values, outliers for different time series can be identified. Outliers are the data points that are failing outside the band as in the case of Fig. 6., and there are two outliers.

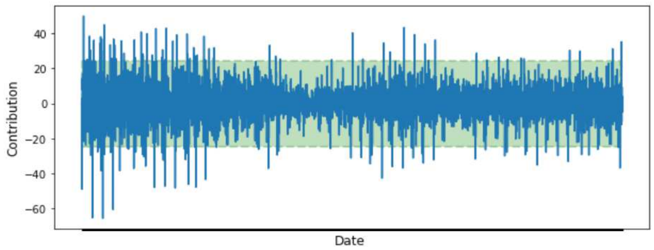

Fig. 6. Residual Band Factor to identify outliers.

Different values of k are used for different locations, and those k values are identified using multi-agent technologies.

IV. Implementation

Since the message space agent technique is proposed as the solution, it is essential to understand the stages of agents' implementation, their task(s), and decisions associated with the relevant multi-agent, as shown in TABLE I.

TABLE I. Different Agents, Agent Types, Tasks, and Decisions

Agent	Agent Type	Task	Decision
Time Series Conversion	Knowledge	Convert State Space data to numerical	Selection of time range
Distance Calculation	Knowledge	Calculate the distance for a given location	Distance between the human and the location
Row Selection	Process	Removal of the location where sufficient data is not available.	
Missing Value Identification	Process	Identification of missing values.	
Missing Value Corrections	Knowledge / Rule	Correcting the missing values either by rule or from the regression.	Best regression model
Outlier Detection	Knowledge	Detecting the outlier human movements.	Best band

The original data set has 112,318,484 records and time series conversion, distance calculation, and row selection were made through multi-agents.

This dataset indicates missing data values as there are different data points for the same time period. Missing value identification is a process agent, and the next task is replacing missing values. As discussed, the regression technique is used to correct the missing values by using attributes such as location, day, hour, month, day of the week, visits at the previous hour, visits at the next hour, visits at next week, visits at the previous week, visits at next week, visits at the previous week, visits at next month and visits at the previous month. TABLE II. shows Root Mean Squared Error values for different regression techniques.

TABLE II. Different Regression Techniques with Root Mean Squared Error

Regression Technique	Root Mean Squared Error
Decision Forest Regression	6.65
Bayesian Linear Regression	6.92
Boosted Decision Tree Regression	**6.11**
Linear Regression	6.92
Neural Network Regression	7.08
Poisson Regression	18.08

Boosted decision tree regression was selected as it has the lowest Root Mean Squared Error value. The selected model is deployed as a web service so that model is consumed very efficiently. After the model is deployed, missing values are predicted.

TABLE III. shows the number of records that were identified from the regression prediction and from the rule.

TABLE III. Combination of Data After Missing value Correction

Location	Record Count			
	Start	Replace by Prediction	Replace by Rule	Total
Hiroshima Castle	8,377	408	-	8,785
Hiroshima Botanical Garden	4,841	3175	729	8,745
Hiroshima Station	8,773	12	-	8,785
Ballpark	8,682	103	-	8,785
Hiroshima Prefectural	8,734	51	-	8,785
Kamiyacho	8,770	15	-	8,785
Hiroshima MOCA	8,530	253	2	
Bomb Site	8,723	61	-	8,785
Hiroshima City Hall	8,785	-	-	8,785
Hachobori, Hiroshima	8,785	-	-	8,785

Except for the two locations (Hiroshima Botanical Garden and Hiroshima MOCA), all the other locations missing data was predicted from the regression model.

Multiple parallel agents were used to replace the missing values to improve performance and scalability. TABLE IV. shows the number of predictions done by each agent and the duration for each agent to process.

TABLE IV. PERFORMANCE OF REGRESSION AGENTS

Agent #	Number of Prediction	Duration (Minutes)
Agent 1	277	7
Agent 2	242	164
Agent 3	222	216
Agent 4	253	14
Agent 5	247	157
Agent 6	218	282
Agent 7	227	194
Agent 8	226	230
Agent 9	262	3
Agent 10	241	58
Agent 11	249	72

After the missing data replacement is completed, stationary testing was carried out. Augmented Dicky-Fuller (ADF) test was done from the process agent and the results are shown in Fig. 7.

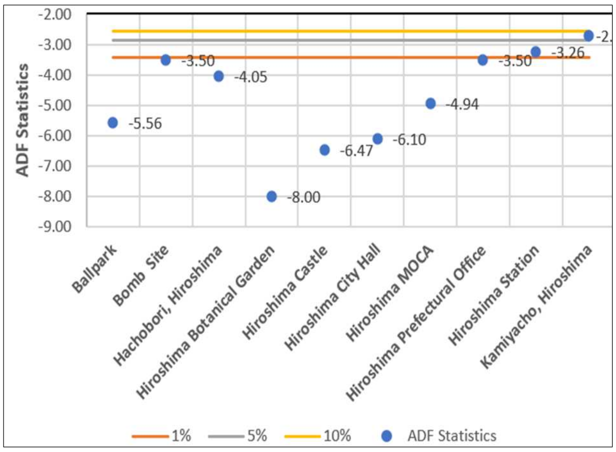

Fig. 7. ADF Statistics for Different Locations vs critical values 1%, 5%, and 10%.

When analysing ADF statistics, the time series of the Kamiyacho, Hiroshima location is non-stationary. This research has ignored the non-stationary locations and outlier detections were carried out for the stationary locations.

Then decomposition of the time series using STL using LOESS was done to detect the outliers in Time Series.

TABLE V. NUMBER OF OUTLIERS FOR DIFFERENT LOCATIONS AND DIFFERENT BAND FACTORS

Location	Band Factor			
	3	4	5	6
Hiroshima Castle	74	27	13	13
Hiroshima Botanical Garden	94	25	6	4
Hiroshima Station	159	54	16	4
Ballpark	105	32	11	1
Hiroshima Prefectural Office	147	50	15	5
Hiroshima MOCA	96	28	12	5
Bomb Site	120	28	5	3
Hiroshima City Hall	93	13	2	1
Hachobori, Hiroshima	144	34	9	0

After the decomposition, the residual dataset was selected, and four different band factors were considered, which are 3,4,5, and 6. TABLE V. shows the number of outliers for different locations for different band factors.

Multi-agents define the selection of the band factor that suits each location. Once the outliers are detected, outliers are plotted in red colour. To further analyse the detected outliers, holidays are plotted differently.

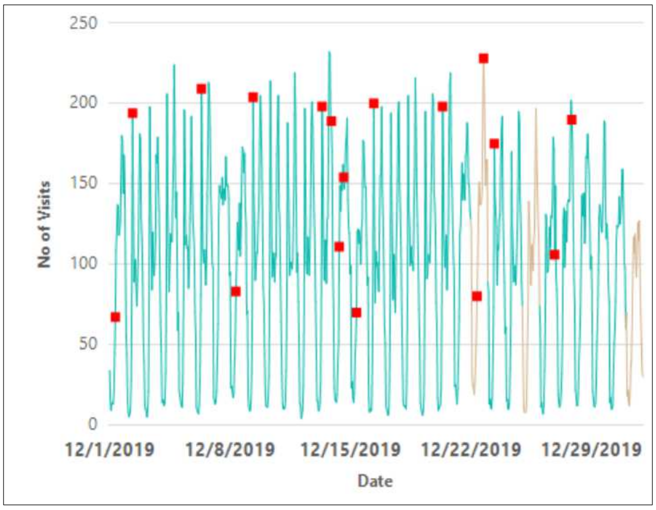

Fig. 8. Unusual movement of humans for a selected location.

Fig. 8. shows the detected outliers for the location of Hiroshima Station. Holidays are indicated in the above graph in order to avoid confusion.

Apart from the unusual movements during the holidays, more outliers can be detected for the location at different hours.

V. CONCLUDING REMARKS

This research has used multi-agent technologies to detect unusual human movement by converting the state-space data to time series data. A mobile dataset in Hiroshima, Japan, was used to prove the proposed concept. Missing value correction was done by the rule and regression technique as time series require a continuous dataset. Multiple self-emerging processing agents were proposed to improve scalability and performance. After the stationary test was done using the ADF and by using the STL technique, outliers were detected. These outliers were plotted with the holidays so that a better analysis could be done. This research can be extended to multiple outlier detection techniques as this research is only limited to one technique.

This research can be extended by fixing the non-stationary issue by including another multi-agent. Further, external factors such as weather data and events can be incorporated into this dataset to predict more accurate outliers.

VI. References

[1] P. Sokannit and P. Chujai, "Forecasting Household Electricity Consumption Using Time Series Models," *International Journal of Mchine Learning and Computing,* vol. 11, no. 6, November 2021.

[2] L. Ji, Y. Zou, K. He and B. Zhu, "Carbon futures price forecasting based with ARIMA-CNN-LSTM Model," in *7th International Conference on Information Technology and Quantitative Management (ITQM 2019)*, 2020.

[3] M. A. Hanjah, "Forecasting Major Fruit Crops Productions in Bangladesh using Box-Jenkins ARIMA Model," *Journal of Economics and Sustainable Development,* vol. 5, no. 7, 2014.

[4] I. E. E. C. Etaga Harrision, "Time Series Models of Crude Oil Production and Export in Nigeria (1999-2015)," *African Journal of Mathematics and Statistics Studies,* vol. 3, no. 1, pp. 1-24, 2020.

[5] A. Belayneh and J. Adamowski, "Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression," *Applied Computational Intelligence and Soft Computing,* vol. 2012, p. 13, 18 July 2012.

[6] A. G. Salman, Y. Heryadi, E. Abdurahman and W. Suparta, "Weather Forecasting Using Merged Long Short-Term Memory Model (LSTM) and Autoregressive Integrated Moving Average (ARIMA) Model," *Science, Journal of Computer,* 26 June 2018.

[7] L. Zhang and X. Peng, "Time Series Estimation of Gas Sensor Baseline Drift using ARMA and Kalman Based Models," *National Natural Science Foundation of China,* vol. 36, no. 1, pp. 34-39, 11 July 2015.

[8] L. Ni, Y. Li, X. Wang, J. Zhang, J. Yu and C. Qi, "Forecasting of Forex Time Series Data Based on Deep Learning," in *2018 International Conference on Identification, Information and Knowledge in the Internet of Things*, China, 2019.

[9] D. A. Dickey. "Stationarity Issues in Time Series Models". Statistics and Data Analysis, North Carolina State University, pp 192.

[10] C. W. Xi Wang, "Time Series Data Cleaning: A Survey," IEEE Access, 21 December 2019.

[11] M. K. Shendea, A. E. Feij'ooLorenzo and Neeraj Dhanraj Bokde, "cleanTS: Automated (AutoML) Tool to Clean Univariate Time Series at Microscales," arXiv, 2021.

[12] Z. Y.C. Liu, S. Roychowdhury, S. Tarlow, A. Nair, S. Badge and T. Shah, "AutoDC: Automated data-centric," Sydney, Australia, 2021

[13] M. V. Muntean and D. Onişa, "Agent for Preprocessing and Forecasting Time-Series Data," 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), pp. 1-4, 2018.

[14] I. Z. S. Raudys, "The Multi-Agent System for Prediction for Financial Time Series," 15th International Conference, ICAISC, 2016.

[15] D. Asanka, "AutoML in Azure Machine Learning for Regression and Time Series," SQLShack, August 2021. [Online]. Available: https://www.sqlshack.com/automl-in-azure-machine-learning-for-regression-and-time-series/. [Accessed 16 August 2021].

[16] R. Patil, "Automate Time Series Forecasting using Auto-TS," [Online]. Available: https://www.analyticsvidhya.com/blog/2021/04/automate-time-series-forecasting-using-auto-ts/. [Accessed 10 November 2021].

[17] D. Asanka, "Time Series Anomaly Detection in Azure Machine Learning," SQLShack, 1 April 2021. [Online]. Available: https://www.sqlshack.com/time-series-anomaly-detection-in-azure-machine-learning/.

[18] S. S. Ho and H. Wechsler, "A Martingale Framework for Detecting Changes in Data Streams by Testing Exchangeability," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 12, pp. 2113-2127, Dec. 2010, doi: 10.1109/TPAMI.2010.48.

[19] J. Lee, K. Jeong, W. Kim, Multivariate time series traffic anomaly detection with Prediction and AutoEncoder. Research Square; 2022. DOI: 10.21203/rs.3.rs-1740184/v1.

[20] P. D. Asanka and A. S. Karunananda, "Troubleshooting in Software as a Service (SaaS) Environments using Multi-Agent Technology," International Journal of Innovative Research in Technology, vol. 8, p. 1, 2014.

[21] H. P. Nii, "The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures," AI Magazine, vol. 7, no. 2, 1986.

[22] R. B. Cleveland, W. S. Cleveland, J. E. McRae, & I. J. Terpenning, (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–33.