Computing and Informatics, Vol. 43, 2024, 1137–1160, doi: 10.31577/cai_2024_5_1137

BAYESIAN INFORMATION CRITERION ANALYSIS FOR ACCURACY IMPROVEMENT OF MULTIVARIATE TIME SERIES DATA ANALYSIS

Ayako Ohshiro

Faculty of Economics
Okinawa International University
Okinawa 901-2701, Japan
e-mail: a.ohshiro@okiu.ac.jp

Morikazu Nakamura

Faculty of Engineering
University of the Ryukyus
Okinawa 903-0213, Japan
e-mail: morikazu@ie.u-ryukyu.ac.jp

Abstract. Multivariate time series data can be collected and employed in various fields to predict future data. However, owing to significant uncertainty and noise, controlling the prediction accuracy during practical applications remains challenging. Therefore, this study examines the Bayesian information criterion (BIC) as an evaluation metric for prediction models and analyzes its changes by varying the explanatory variables, variable pairs, and learning and validation periods. Descriptive statistics and decision tree-based algorithms, such as classification and regression tree, random forest, and dynamic time warping, were employed in the analysis. The experimental evaluations were conducted using two types of restaurant data: sales, weather, number of customers, number of views on gourmet site, and day of the week. Based on the experimental results, we compared and discussed the learning behavior based on various explanatory variable combinations. We discovered that 1. the explanatory variable, the number of customers, exhibited a significantly different trend from other variables when dynamic time warping was applied, particularly in combination with other variables, and 2. variables with seasonality yielded the best performance when used independently; otherwise, the predictive accuracy decreased according to the decision tree results. This comparative investigation revealed that the proposed BIC analysis method proposed can be used to effectively identify the optimal combination of explanatory variables for multivariate time series data that exhibit characteristics such as seasonality.

Keywords: Multivariate time series analysis, Bayesian information criterion, decision tree, dynamic time warping

1 INTRODUCTION

Owing to the advancements in data acquisition and collection technologies, data-driven strategic analysis and forecasting have been widely adopted across various industries. Additionally, both the government and private sectors recognize that data-oriented strategic planning is essential for appropriately executing tasks.

As for econometric time-series analyses, several fundamental models have been proposed, including auto regression (AR), moving average (MA), auto regression integrated moving average (ARIMA) [1], and generalized autoregressive conditional heteroscedasticity (GARCH), which is based on fluctuation prediction [2].

In addition to the aforementioned traditional statistical analysis approaches, machine learning-based methods have been gaining popularity in time-series data research. For instance, recurrent neural networks (RNNs) and long short-term memory (LSTM) models are frequently employed for time-series data prediction. However, the selection of prediction models for time series is competitive because the prediction performance can be highly case-specific. Furthermore, the prediction performance depends on several factors, including the nature of the data, diverse preprocessing techniques, feature selection, and algorithm selection.

For instance, a study [3] that compared the performance of various models in a time-series analysis demonstrated that the ARIMA model [4], originally proposed to handle business or economic data, is substantially accurate compared to advanced techniques such as deep learning and LSTM. Moreover, the authors of [5] indicated that the accuracy of the ARIMA model is equivalent to that of deep learning approaches and can even be extended to multivariate analyses.

Several hybrid approaches have also been proposed; for instance, a combination of the ARIMA model and deep learning has been proposed for time-series analysis [6], and deep learning techniques have been introduced to learn the characteristics of time series data [7, 8, 9]. We also examined an approach [10] that emphasizes adjusting seasonality in time series data using a proportional-integral-derivative (PID) control mechanism. Additionally, accuracy has been further improved by incorporating Bayesian networks [11]. Moreover, dynamic regression models with Bayesian networks for time series data have been researched, as demonstrated in [12].

To predict sightseeing demand, a neural-network-based time-series analysis has been proposed [13], and a study comparing neural networks and traditional time-series analysis for predicting single-variable time series data has also been published [14]. Yuyama et al. [15] introduced a method to enhance the accuracy of wind-speed predictions. This technique is based on analyzing prediction errors using machine learning algorithms for the time series of wind-speed data. Notably, the identified error tendencies were successfully leveraged as a feature to augment prediction accuracy. Furthermore, some studies have employed time-series analyses in the medical field, such as forecasting dengue hemorrhagic fever cases [16] and estimating the number of beds required during a SARS outbreak [17].

Owing to the increasing ease of obtaining diverse time series data and the growing practice of combining them into multivariate sets, the application of time series data analysis has been extended to numerous fields, including medicine [18, 19, 20]. Consequently, multiple regression analysis based on multivariate data have drawn more attention than univariate analyses. As variable selection is crucial for developing a multiple regression model, various methods have been proposed, such as the stepwise approach. However, it is important to note that although the predictive accuracy is likely to improve with more variables, it often fluctuates based on the training data and variable combination employed.

This study aimed to advance the field of multivariate regression time-series analysis by preemptively estimating the optimal combination of variables and understanding the tendency of the accuracy estimated for the prediction model. We focused on enhancing the accuracy of a sales prediction model created using time series data and present our findings based on a fluctuation analysis of the Bayesian information criterion (BIC) over the time-series transition, particularly emphasizing on the BIC characteristics for the model. Furthermore, we delineated the distinct characteristics of each variable during model generation and the resultant shifts in the BIC. In particular, the main purpose of our research is to observe the behavior due to the changes in combinations of variables based on our proposed method, for obtaining datasets. Thus, the accuracy of proposal forecasting has not been compared with existing methods. The experimental data employed in this study comprised restaurant sales figures of a popular tourist destination in Japan, along with related data.

The remainder of this paper is organized as follows. Section 2 provides a succinct background of the BIC and time-series analysis. Section 3 introduces the proposed BIC analysis-based variable selection method for multivariate time series data. Section 4 discusses the experimental results. Finally, Section 5 concludes the paper and proposes directions for future work.

2 METHODS

This section provides a brief overview of certain time-series analytics and evaluation indices that are frequently used in this study.

Dynamic Time Warping. In the following sections, the evaluation indices and algorithms implemented using the proposed method are explained. The dynamic time warping (DTW) technique is used to calculate the similarity between multiple time-series datasets, which allows establishing the shortest path for the time-series data. This path is derived by calculating the distance between each point in the two time-series datasets using a brute-force approach:

$$\mathrm{DTW}(x, y) = \min \sum_{t=1}^{p} |x_t - y_t|, \tag{1}$$

where $x = (x_1, x_2, \ldots, x_p)$ and $y = (y_1, y_2, \ldots, y_p)$ denote the two time-series datasets.

DTW is commonly employed for human behavior analyses, including voice [21] and locomotion data. Furthermore, the application of DTW to cluster time series data has been demonstrated in various domains, such as education [22]. Additionally, it has been used in studies focused on detecting student grit using a clustering method [23].

Bayesian Information Criterion. Several methods are available for evaluating the performance of prediction models. These include the root mean squared error (RMSE); R^2 (the coefficient of determination); Akaike information criterion (AIC) [24], which considers the number of parameters and the model likelihood; and BIC [25], which additionally considers the number of data points.

BIC is particularly useful for model selection, and is expressed as follows:

$$\mathrm{BIC} = -2\log(L) + k\log(n), \tag{2}$$

where L, n, and k denote the likelihood function, number of data points, and number of parameters, respectively.

Numerous studies have investigated model validation using various types of data processing methods in conjunction with both the AIC and BIC [26, 27].

The use cases of time series data have expanded beyond traditional applications in economics and meteorology to many other fields, including medicine [18, 19, 20]. This trend has highlighted the need for multiple regression analysis using multivariate datasets as opposed to univariate data. It is worth noting that predictive accuracy is likely to improve with a higher number of variables. However, this improvement often varies depending on the training data and variable combination used.

For a multivariate regression time series analysis, predicting the trend of the estimation accuracy of the constructed model and pre-determining the most effective combination of variables can significantly contribute to the construction of predictive models with optimal variable combinations. This study proposes a method to improve the accuracy of sales prediction models generated using time series data. It focuses on a fluctuation analysis of BIC in a time-series transition, with a particular focus on the characteristics of the model's BIC. This study presents the process and performance results of the proposed analytical method in detail.

Additionally, we describe the characteristics of each variable used for model generation and the factors behind changes in BIC. We used variables related to restaurant sales of a popular tourist location in Japan as the experimental data.

Dynamic Regression Models. Time series analysis typically forecasts future values by leveraging past changes in univariate data. However, it can also accommodate multivariate data by using dynamic regression models [28]. In the general regression model, assuming a predefined linear relationship between y_t and x_t at time t, y_t can be expressed using x_t, β_i (representing the gradient), and ϵ_t (the error term denoting the deviation from the regression function) as follows:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t. \tag{3}$$

By introducing autocorrelation as η and based on the ARIMA process, the following equation is obtained:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots \beta_k x_{k,t} + \eta_t. \tag{4}$$

Moreover, the optimal combination of variables is determined based on criteria such as the AIC, BIC, and lag value. The objective is to minimize the values of both ϵ_t and η_t.

3 BIC ANALYSIS-BASED VARIABLE SELECTION FOR MULTIVARIATE TIME SERIES DATA

In this section, we propose a BIC analysis-based algorithm for selecting variables from multivariate time series data.

Prediction models for time series data can be constructed by combining multiple variables through a multiple regression analysis. Variable selection significantly influences both the prediction process and results. The characteristics of the variables can be determined by observing the changes in BIC based on the selection of different variables.

Cross-validation is commonly employed to ensure robust validation during time series analysis. Typically, the first m steps in the time series data are employed as training data, whereas the subsequent n steps are used as test data. Thereafter, the periods are iteratively shifted forward. In this context, the BIC may fluctuate according to the period selected. This study aims to investigate model accuracy by analyzing BIC variations and using the results to guide variable selection.

The analysis of these changes can include various aspects such as statistical features, time-series changes, and feature classification. We employed descriptive statistics for statistical features, DTW for time-series changes, and decision tree models for feature classification.

The proposed method for improving the accuracy of multivariate time series data analysis through BIC analysis comprises the following steps:

Step 1. Define the training and test periods for multivariate time series data.

Step 2. Enumerate the combinations of variables for multivariate time series data.

Step 3. Calculate the BIC for each combination of variables using a dynamic regression model.

Step 4. Observe the effects of the training period variations on BIC.

Step 5. Analyze the effects of variable combinations on BIC by using descriptive statistical values for each test period.

Step 6. Apply DTW to each variable combination and observe the effects of variable combinations at the time-series level.

Step 7. Analyze the relationship between BIC fluctuations and each variable combination by applying a decision tree-based classification algorithm.

Figure 1 illustrates the process of BIC analysis-based variable selection for multivariate time series data. Steps 1–3 pertain to data generation using the dynamic regression model and cross-validation for BIC analysis. Steps 4 and 5 involve the application of descriptive statistics. Step 6 corresponds to the analysis of time series data using DTW, whereas Step 7 involves feature classification using decision tree models.

Furthermore, to discuss the variations in results owing to the combination of multiple variables, the following three hypotheses were formulated and tested using the proposed analytical method:

H_0**:** BIC prediction accuracy improves as the number of variable combinations increased.

H_1**:** BIC prediction accuracy changes regardless of the number of variable combinations employed.

H_2**:** BIC prediction accuracy changes based on the characteristics of the variables.

4 RESULTS AND DISCUSSION

In this section, we present the experimental evaluation and discuss the results. First, we report the results of the BIC analysis for the dynamic regression models based on the explanatory variable combinations employed, as described in Step 5 of the proposed procedure. Thereafter, we present the results of the BIC classification of time series data using DTW, as explained in Step 6. Finally, we discuss the analysis results of the contributing variables and their combinations for decision tree algorithms, as detailed in Step 7.

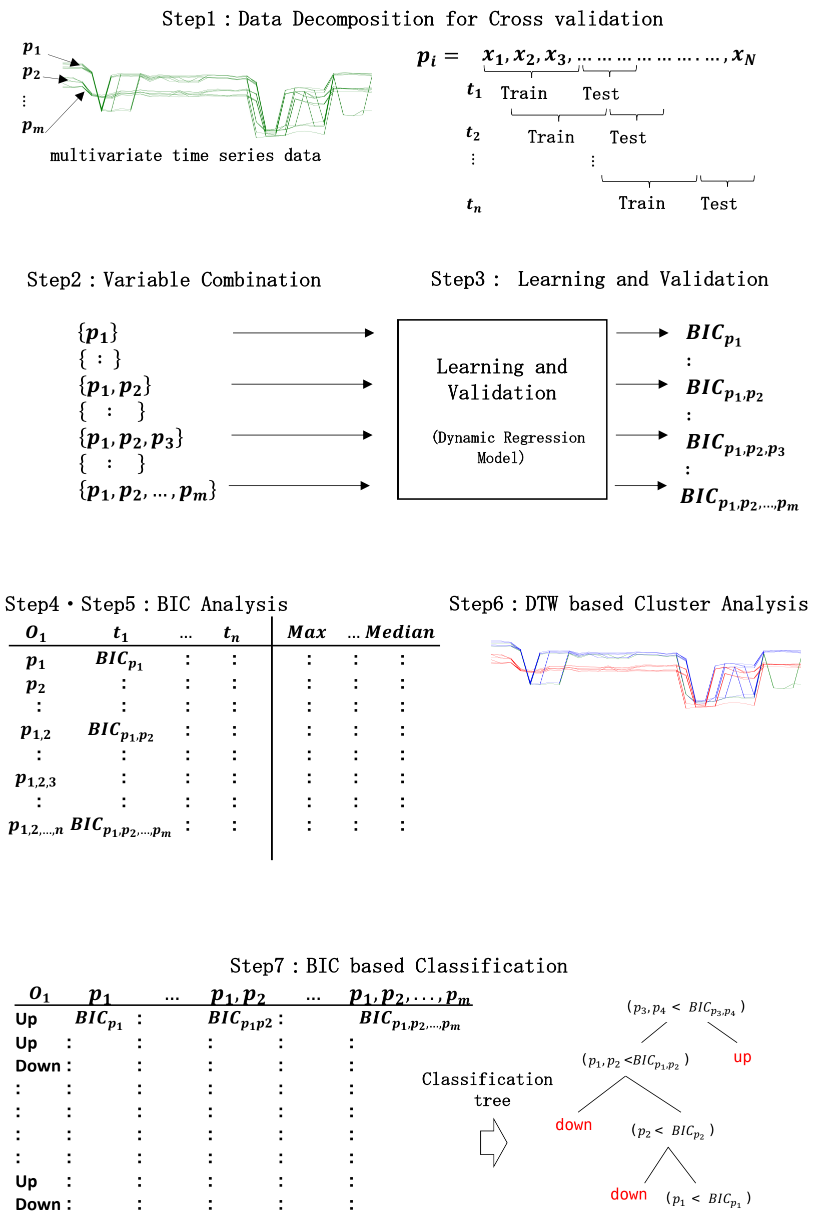


Figure 1. Process of BIC analysis-based variable selection for multivariate time series data

4.1 BIC Analysis Based on Variable Combination

As previously described, prediction models can be constructed using multiple regression analyses by combining multiple variables, whose characteristics can be determined by observing the changes in BIC according to various selections. For the experimental data, we used actual daily data collected from a restaurant in Japan, with a focus on total sales. Therefore, we defined the periods for the training and validation data as follows:

- Length of training data: 20 weeks;
- Length of validation data: 5 weeks.

The variables used in this study and their combinations used for model generation are listed in Tables 1 and 2, respectively. As shown in Table 1, we could use at most five explanatory variables at a time. Table 2 lists the combinations of variables selected. For example, $Case_{1,2}$ indicates the selection of the number of customers and the probability of precipitation variables.

Variable	Description
O_t	Total sales
p_1	Number of customers
p_2	Probability of precipitation
p_3	Number of views on a gourmet site
p_4	Day of the week

Table 1. Available variables

Notation	Variable Combination
$Case_1$	$\{p_1\}$
$Case_2$	$\{p_2\}$
$Case_3$	$\{p_3\}$
$Case_4$	$\{p_4\}$
$Case_{1,2}$	$\{p_1, p_2\}$
$Case_{1,3}$	$\{p_1, p_3\}$
$Case_{1,4}$	$\{p_1, p_4\}$
$Case_{2,3}$	$\{p_2, p_3\}$
$Case_{2,4}$	$\{p_2, p_4\}$
$Case_{3,4}$	$\{p_3, p_4\}$
$Case_{1,2,3}$	$\{p_1, p_2, p_3\}$
$Case_{1,2,4}$	$\{p_1, p_2, p_4\}$
$Case_{1,3,4}$	$\{p_1, p_3, p_4\}$
$Case_{2,3,4}$	$\{p_2, p_3, p_4\}$
$Case_{1,2,3,4}$	$\{p_1, p_2, p_3, p_4\}$

Table 2. Combinations of variables

Multivariate regression analysis was performed with seasonality set to 1 week (seasonality = 7). The regression coefficients and BIC varied depending on the training data period and the number or combination of variables. Table 3 presents the BIC fluctuations for predicting total sales when the periods of the training data, from t_1–t_n, and the combination of explanatory variables, from $Case_1$ to $Case_{1,2,3,4}$, are altered. The total sales from restaurants (denoted as D_a) were used as experimental data. To conserve space, we used the notation C_x for $Case_x$ in the first column of the table. Max, Min, Median, and SD denote the maximum, minimum, median, and standard deviation of BIC, respectively.

The regression coefficients and BIC varied depending on the training data period and combination of variables employed. Table 3 presents the changes in BIC for predicting total sales using various explanatory variables and varying the training data period from $t = 1$ to $t = 69$. For the experimental data, we used the total sales from restaurants (represented as D_a). To save space, we used the notation C_x for $Case_x$ in the first column. Max, Min, Median, and SD denote the maximum, minimum, median, and standard deviation of BIC, respectively.

	$t = 1$	$t = 2$ ...	$t = 68$	$t = 69$	Max	Min	Median	SD
O_t	4 067.00	4 069.13	4 098.16	4 093.31	4 136.91	3 807.42	3 859.67	132.83
C_1	4 008.13	4 026.57	$\underline{4\,048.82}$	4 043.12	$\underline{4\,090.98}$	$\underline{3\,797.74}$	$\underline{4\,011.97}$	$\underline{56.98}$
C_2	4 067.44	4 068.81	4 100.06	4 092.83	$\underline{4\,142.61}$	$\underline{3\,807.93}$	3 863.12	$\underline{134.87}$
C_3	4 059.55	4 057.54	4 097.40	4 090.78	$\underline{4\,141.62}$	$\underline{3\,796.04}$	4 053.69	102.21
C_4	$\underline{3\,847.15}$	4 032.93	4 102.93	4 098.18	4 110.89	3 788.49	4 023.07	123.03
$C_{1,2}$	3 974.97	4 014.94	4 050.82	4 040.48	4 075.14 $\downarrow$	3 782.35 $\downarrow$	$\underline{3\,997.95}$	$\underline{41.94}$ $\downarrow$
$C_{1,3}$	3 980.36	4 000.54	4 053.47	4 047.36	4 062.64 $\downarrow$	3 780.71 $\downarrow$	3 999.43 $\downarrow$	$\underline{54.12}$ $\downarrow$
$C_{1,4}$	4 036.17	4 037.62	4 102.15	4 095.70	4 109.28	3 789.71	4 038.78	120.41
$C_{2,3}$	3 991.47	4 027.77	4 056.12	4 045.37	4 092.76	3 801.10	4 016.34	51.86 $\downarrow$
$C_{2,4}$	4 062.16	4 058.64	4 099.20	4 088.91	4 147.59 $\uparrow$	3 796.31	4 053.28	105.39 $\downarrow$
$C_{3,4}$	4 044.65	4 034.11	4 104.51	4 097.66	4 115.80	3 790.31	4 041.72	129.28 $\uparrow$
$C_{1,2,3}$	3 951.51	$\underline{3\,996.6}$	4 055.44	4 044.92	$\underline{4\,061.91}$ $\downarrow$	$\underline{3\,773.24}$ $\downarrow$	3 993.66	61.80
$C_{1,2,4}$	4 037.22	4 039.95	4 103.59	4 093.8	4 112.80	3 791.95	4 042.31	124.14
$C_{1,3,4}$	3 985.02	4 016.63	4 054.16	$\underline{4\,042.81}$	4 083.70 $\downarrow$	3 783.55 $\downarrow$	4 002.83 $\downarrow$	42.65 $\downarrow$
$C_{2,3,4}$	3 967.67	4 002.93	4 060.48	4 049.71	4 077.47 $\downarrow$	3 784.57 $\downarrow$	4 005.98	54.81 $\downarrow$
$C_{1,2,3,4}$	3 953.36	3 999.54	4 058.15	4 047.37	4 066.67	3 775.75 $\downarrow$	3 997.13	$\underline{57.79}$

Table 3. Changes in BIC and descriptive statistics according to the selection of explanatory variables for D_a

Tables 4 a) and 4 b) present the correlation matrices of all variables. As the patterns of the single variables, the correlation coefficients for the combination of C_1 and C_4 are the smallest. According to the result of D_a, the correlation coefficients of C_2 tends to be larger than those of other variables, whereas those of C_4 tends to be smaller. As the patterns of the single variables, the correlation coefficients for the combination of C_1 and C_2 are the smallest. Furthermore, the correlation coefficients of $C_{2,4}$ tends to be small, which is a common feature based on the results of D_a and D_b.

To investigate more characteristics of multivariate time series data, we extracted coefficient of variance and gradient. To obtain a "Gradient", the change in BIC was converted into three values and defined as follows:

$$d_i = \begin{cases} -1, & t_i < t_{i-1}, \\ 0, & t_i = t_{i-1}, \\ 1, & t_i > t_{i-1}, \end{cases} \tag{5}$$

$$\mathit{Gradient} = \frac{|\{d_i | d_i = 1\}|}{|n|}. \tag{6}$$

Table 5 presents the coefficients of variance and gradient for all variables of datasets. The coefficient of variance indicates the fluctuation and is one of the evaluation indices used to measure changes in time series data. The coefficient of variance of O_t indicates "Total sales" and that of C_2 indicates "Probability of precipitation" for D_A; D_B is over 0.03 and the variation is large compared to other variables. The *Gradient* of $C_{1,2,4}$ for D_A and $C_{1,3,4}$ for D_B are large, and the extent of increase is larger than that of other variables. Furthermore, as the maximum of the *Gradient* for D_A is less than 0.3, it can be predicted that the number of decrease time than the number of increase time about the time series data for D_A.

First, we analyze the variations in the model accuracy with respect to different training periods. For instance, when the training period was $t = t_1$, the regression model yielding the highest estimation accuracy was $Case_4$, which included the Day of the Week variable. By contrast, when the training period was $t = t_2$, the regression model with the highest estimation accuracy was $Case_{1,2,3}$, which included the Number of Customers, Probability of Precipitation, and Number of Views on Gourmet Site variables. This implies that the combination of variables minimizes BIC changes according to the learning period. By forecasting BIC behavior and identifying the most effective variable combination for model generation, we can optimize store sales prediction accuracy.

Next, we examine the descriptive statistics of model accuracy. $\downarrow$ and $\uparrow$ denote increase and decrease in BIC values, respectively, compared with the single-variable case. When focusing on the maximum and minimum values, the BIC of multivariable cases, such as $Case_{1,2}$, $Case_{1,3}$, and $Case_{1,2,3}$, decreased compared to that of simpler cases, such as $Case_1$, $Case_2$, and $Case_3$. This suggests that the prediction model accuracy is more likely to improve by combining multiple variables rather than using each variable individually. Similarly, regarding the SD, the BIC variation in multivariable cases, such as $Case_{1,2}$ and $Case_{1,3}$, decreased compared with that in simpler cases, indicating that the model accuracy was improved. Furthermore, $Case_{1,2}$ and $Case_{1,3}$ exhibited better performance than the all-variable case $Case_{1,2,3,4}$.

a) Correlation for D_a

	C_1	C_2	C_3	C_4	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{2,3}$	$C_{2,4}$	$C_{3,4}$	$C_{1,2,3}$	$C_{1,2,4}$	$C_{1,3,4}$	$C_{2,3,4}$	$C_{1,2,3,4}$
O_t	0.5996	0.9830	0.6875	0.6920	0.6240	0.7150	0.5900	0.7010	0.8029	0.7254	0.7225	0.5909	0.5889	0.7298	0.6019
C_1		0.6029	0.6969	0.4874	0.9180	0.7282	0.9057	0.7123	0.5467	0.6030	0.7354	0.9154	0.9783	0.5973	0.8923
C_2			0.6877	0.6762	0.6307	0.7354	0.5994	0.7056	0.7924	0.7220	0.7416	0.6015	0.5996	0.7298	0.6168
C_3				0.6786	0.6797	0.6808	0.6701	0.8696	0.6954	0.7941	0.6772	0.6813	0.7014	0.7665	0.6794
C_4					0.4875	0.4954	0.4808	0.5732	0.8843	0.7715	0.5044	0.4884	0.4849	0.7025	0.4791
$C_{1,2}$						0.7647	0.7944	0.7008	0.5561	0.5997	0.7753	0.8079	0.9025	0.5977	0.9694
$C_{1,3}$							0.7742	0.6798	0.5903	0.6361	0.9954	0.7776	0.7610	0.6280	0.7938
$C_{1,4}$								0.7113	0.5660	0.6257	0.7819	0.9961	0.9104	0.6330	0.8020
$C_{2,3}$									0.6867	0.7281	0.6750	0.7244	0.7109	0.7993	0.6977
$C_{2,4}$										0.8562	0.5954	0.5763	0.5519	0.8368	0.5604
$C_{3,4}$											0.6417	0.6418	0.6388	0.9338	0.6384
$C_{1,2,3}$												0.7848	0.7670	0.6320	0.8016
$C_{1,2,4}$													0.9213	0.6498	0.8184
$C_{1,3,4}$														0.6351	0.9253
$C_{2,3,4}$															0.6407

b) Correlation for D_b

	C_1	C_2	C_3	C_4	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{2,3}$	$C_{2,4}$	$C_{3,4}$	$C_{1,2,3}$	$C_{1,2,4}$	$C_{1,3,4}$	$C_{2,3,4}$	$C_{1,2,3,4}$
O_t	0.7014	0.8239	0.7525	0.7669	0.8989	0.6478	0.6368	0.7843	0.6282	0.7636	0.7656	0.7476	0.6067	0.8340	0.7444
C_1		0.5622	0.8133	0.6707	0.7416	0.8658	0.8541	0.7188	0.4103	0.7683	0.7694	0.7611	0.8434	0.5857	0.7508
C_2			0.6360	0.6337	0.7345	0.5180	0.4701	0.6641	0.8291	0.6268	0.6216	0.5752	0.4526	0.7069	0.5970
C_3				0.8938	0.7367	0.9150	0.8870	0.9292	0.6672	0.9218	0.8586	0.8407	0.8832	0.8328	0.8401
C_4					0.7843	0.7845	0.7581	0.9468	0.6906	0.8213	0.8930	0.8891	0.7564	0.9045	0.8899
$C_{1,2}$						0.6582	0.6240	0.7899	0.6796	0.6627	0.8155	0.7873	0.5928	0.8689	0.7956
$C_{1,3}$							0.9814	0.8156	0.5107	0.8638	0.8547	0.8556	0.9871	0.6899	0.8475
$C_{1,4}$								0.7817	0.4463	0.8654	0.8157	0.8403	0.9939	0.6520	0.8225
$C_{2,3}$									0.7001	0.8517	0.9478	0.9349	0.7761	0.8904	0.9405
$C_{2,4}$										0.5746	0.6034	0.5680	0.4343	0.7718	0.5965
$C_{3,4}$											0.7900	0.7951	0.8598	0.7604	0.7780
$C_{1,2,3}$												0.9861	0.8141	0.8206	0.9916
$C_{1,2,4}$													0.8347	0.8048	0.9919
$C_{1,3,4}$														0.6391	0.8194
$C_{2,3,4}$															0.8160

Table 4. Correlation between variables

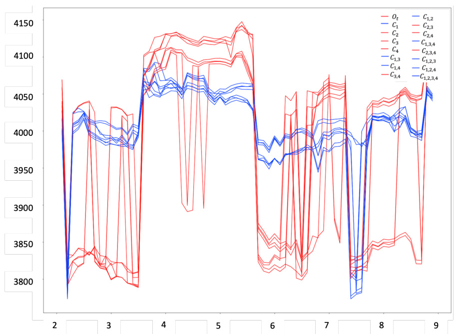
	D_A		D_B	
	CV	*Gradient*	*CV*	*Gradient*
O_t	$\underline{0.033835746}$	0.2319	$\underline{0.030835325}$	0.3889
C_1	0.014822804	0.2174	0.017515224	0.3611
C_2	$\underline{0.034247191}$	0.2464	$\underline{0.031309082}$	0.4444
C_3	0.02698482	0.2319	0.019616495	0.4167
C_4	$\underline{0.030722214}$	0.2609	0.023302834	0.4167
$C_{1,2}$	0.013481039	0.2174	0.025825251	0.4444
$C_{1,3}$	0.010596037	0.2029	0.014370442	0.4167
$C_{1,4}$	0.013807141	0.2464	0.014144489	0.4722
$C_{2,3}$	0.027729362	0.2464	0.022399088	0.3889
$C_{2,4}$	$\underline{0.032738189}$	$\underline{0.2609}$	0.029888159	0.4722
$C_{3,4}$	$\underline{0.030609652}$	0.2464	0.020274866	0.4167
$C_{1,2,3}$	0.010803486	0.2174	0.018425341	0.4167
$C_{1,2,4}$	0.014022714	$\underline{0.2754}$	0.017128383	0.4167
$C_{1,3,4}$	0.015846182	0.2609	0.013657935	$\underline{0.5278}$
$C_{2,3,4}$	$\underline{0.031630946}$	0.2464	0.027224456	0.4444
$C_{1,2,3,4}$	0.014608557	0.2609	0.016973267	0.4722

Table 5. Coefficient of variance (CV) and *Gradient* for all variables of all datasets

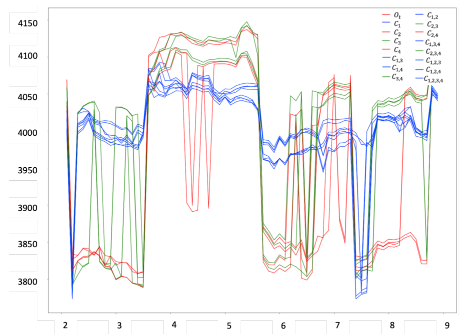
Generally, the accuracy of a prediction model in a multiple regression analysis improves as the number of explanatory variables increases. In this study, the selection of multiple explanatory variables lowered the variations in BIC than using a single variable. However, if an appropriate combination of variables is pre-selected, a regression model with reliable accuracy can be obtained without increasing the number of explanatory variables. Therefore, the variable combination is crucial. If BIC changes can be predicted, an appropriate set of explanatory variables can be selected during the generation of the regression model. In the following section, each combination of variables is visualized as time series data, and their trends are observed by classifying the fluctuations using DTW.

4.2 Clustering of Time Series Data with DTW

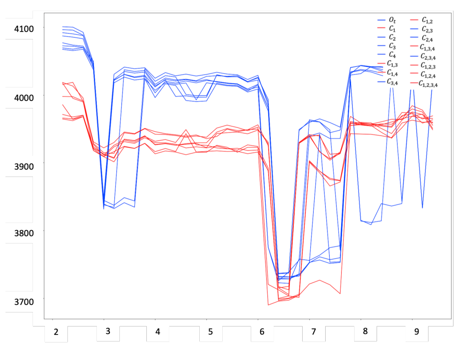
In the previous section, we examined the impact of various variables and their combinations by observing the descriptive statistical value of BIC. In this section, we focus on the transitions in single and multivariable cases and examine the trends of BIC changes. For the experimental data, we used the total sales from the two restaurants (denoted as D_a and D_b), in addition to the data used in the previous section. The results of applying DTW to multiple time series data, as described in Table 2, are presented. We attempted extract features of the time series data from the classification results using DTW in combination with the k-means algorithm. The classification results for cases with two and three clusters for both experimental datasets (D_a and D_b) are shown in Figures 2 a)–2 d).



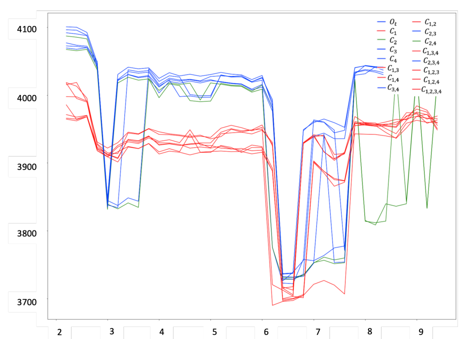
a) Clustering for D_a (number of clusters = 2)



b) Clustering for D_a (number of clusters = 3)



c) Clustering for D_b (number of clusters = 2)



d) Clustering for D_b (number of clusters = 3)

Figure 2. Clustering Results for Time Series Data

We analyzed the trends in BIC changes for each dataset in scenarios with different numbers of clusters. In case of two clusters, the data for D_a were classified into a red line (indicating cluster 0) and a blue line (indicating cluster 1), as shown in Figure 2 a). Among the five patterns of single variables, O_t and $Case_1$–$Case_4$, $Case_1$ (number of customers) were classified as Cluster 1, whereas the remaining variables were classified as Cluster 0. This suggests that variables other than that classified as $Case_1$ exhibited similar trends. Regarding the six patterns involving combinations of two variables, $Case_{3,4}$, $Case_{2,3}$, and $Case_{2,4}$ were classified as Cluster 0, whereas the other combinations were classified as Cluster 1. This means that the combinations including $Case_1$ were categorized separately from other combinations. We observed similar results for the combination of three variables among the four patterns and the combination of four variables: combinations including $Case_1$ were classified as Cluster 1, whereas the other combinations were classified as Cluster 0. This indicates that $Case_1$ significantly affected the other variables. When the number of clusters was increased to three, the data for D_a were indicated as red (Cluster 0), blue (Cluster 1), and green (Cluster 2) lines, as shown in Figure 2 b).

Among the five patterns of single variables, $Case_1$ was classified as Cluster 1, $Case_3$ as Cluster 2, and the remaining as Cluster 0.

Next, we discuss the results of D_b. In the case of two clusters, $Case_1$ was classified as Cluster 1 and the other variables as Cluster 0 for single variables, as shown in Figure 2 c). Furthermore, for combinations of two, three, and four variables, the classification was similar to that of D_a, with combinations that included $Case_1$ separated from other combinations. In case of three clusters, $Case_1$ was classified as Cluster 1, $Case_2$ as Cluster 2, and the other variables as Cluster 0 for single variables, as shown in Figure 2 d). For time series data with combinations of two, three, and four variables, the cases with variable combinations including either $Case_1$ or $Case_2$ were classified into the same cluster as the corresponding case. In the case of variable combinations, both $Case_1$ and $Case_2$ were classified into the same cluster as $Case_1$.

In summary, we demonstrated that $Case_1$ has a significant feature that also affects the results when used in combination with multiple variables based on the comparative classification results obtained by applying DTW.

4.3 BIC-Based Analysis of Effective Variable Combination

In the preceding section, we classified the characteristics of each variable combination by treating the fluctuations as time series data. Based on these results, we discussed the effects of integrating multiple variables on analysis accuracy and the interrelationships among the variables. In this section, we more comprehensively investigate the relationship between the analysis accuracy and variable combinations.

Sale change forecasting is a critical component of restaurant management. In this study, we employed a decision tree to identify the influential explanatory variables for predicting changes in BIC based on total sales. The specific procedure is as follows:

1. Transform the changes in BIC into binary data, such as by contrasting "increase/decrease" relative to the preceding month as "Up/Down", and employ this as the target variable.
2. Define the explanatory variables as the BIC values obtained when employing the variable combinations presented in Table 2.
3. Apply the random forest algorithm to the data synthesized in Steps 1 and 2, and extract the effective combinations of variables and the combinations conducive to predicting changes in BIC.

We identified the crucial variables for predicting BIC transitions to forecast store sales using decision tree algorithms, such as random forest [29] and CART [30]. Decision tree algorithms highlight the essential variables for generating decision rules by employing Gini coefficients or variable importance. Similar to the previous section, we used total sales from the two restaurants (denoted as D_a and D_b) as experimental data. Tables 6 and 7 present the results of sorting Gini coefficients in descending order and the accuracy of the decision rule for D_a, respectively, when the random forest algorithm was applied.

Variable Combination	Mean Decreased Values of Gini
$Case_{1,3}$	3.207450
$Case_{3}$	2.954016
$Case_{2}$	2.944125
$Case_{2,3}$	2.691539
$Case_{1,2,3}$	2.365642
$Case_{4}$	2.352224
$Case_{1,2,4}$	2.248488
$Case_{1,3,4}$	2.189068
$Case_{3,4}$	2.164238
$Case_{2,3,4}$	2.156108
$Case_{1}$	2.135660
$Case_{2,4}$	2.111797
$Case_{1,2}$	2.082316
$Case_{1,4}$	2.011359

Table 6. Random forest: Gini coefficient of each variable for D_a

	Down	Up	class.error
Down	21	17	0.4473684
Up	17	14	0.5483871
Total	38	31	0.4928

Table 7. Random forest: Discriminant results of rules for D_a

From the random forest results, 13 variable combinations were identified as contributing variables. As evident from Table 6, three variable combinations exhibited high Gini coefficients for BIC changes: "Number of Customers" + "Number of Views on a Gourmet Site", "Number of Views on a Gourmet site", and "Probability of Precipitation." $Case_1$, which solely considered the "Number of Customers", had the lowest contribution toward prediction when it was used as the only explanatory variable. However, the performance could be enhanced by combining it with other variables. Conversely, when "Day of the Week" from $Case_4$ was coupled with other variables, the value of the Gini coefficient diminished, suggesting that employing it independently may contribute more effectively toward BIC change prediction. Because $Case_4$ exhibited seasonality, which was distinct from the other variables, it was advantageous to use it individually. Because some cases with fewer explanatory variables outperformed cases with more variables, it is imperative to carefully consider the combination of explanatory variables. Furthermore, as shown in Table 7, more than 60 % of cases were accurately classified as "down" data, and fewer than 50 % of cases were accurately classified as "up" data, with the total rule accuracy exceeding 50 %. In essence, the number of accurate branches surpassed that of inaccurate ones.

Figure 3 shows a list of decision rules for sales changes.

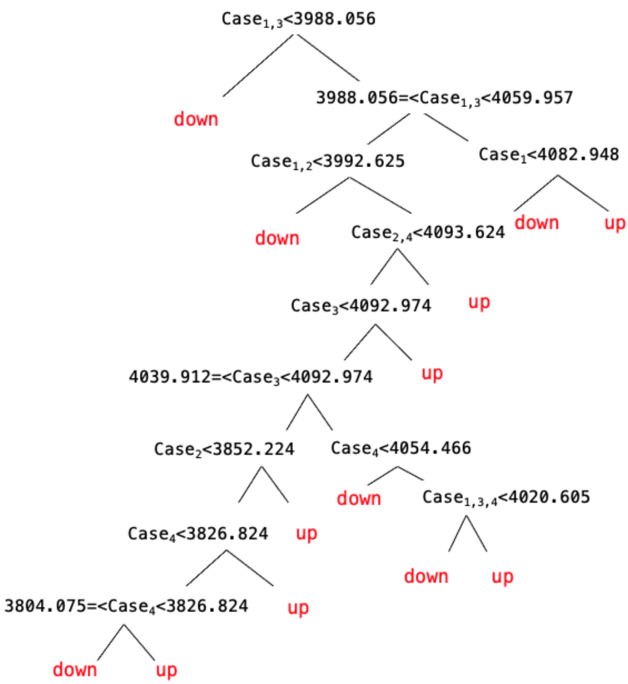


Figure 3. Random forest: Rules acquired for D_a

At the highest branching point, $Case_{1,3}$, which includes the "Number of Customers" and "Number of Views on Gourmet Sites" variables, determined the "down" of BIC for total sales through the combination of these variables. The prediction rules predominantly used $Case_1$, $Case_2$, $Case_3$, $Case_4$, $Case_{1,2}$, $Case_{1,3}$, $Case_{2,4}$, and $Case_{1,3,4}$, which primarily constitute single- or two-variable combinations, except for $Case_{1,3,4}$. Therefore, an appropriate combination of explanatory variables can contribute more to the prediction accuracy than cases with a more substantial number of variables.

Moreover, Table 8 presents the results of the variable importance in a descending order, and Figure 5 displays the discriminant rules for the case of D_a obtained through the CART algorithm.

Based on the results of the CART algorithm, ten variable combinations were identified as contributing variables. First, we discuss the similarities with the results of random forest. The importance of the "Day of the Week" variable in $Case_4$ diminished when it was paired with other variables, suggesting that using it indi-

Variable Combination	Values
$Case_3$	17
$Case_4$	15
$Case_{1,3}$	12
$Case_{2,3}$	11
$Case_{3,4}$	10
$Case_{1,4}$	9
$Case_{2,4}$	9
$Case_2$	6
$Case_1$	5
$Case_{1,2}$	4

Table 8. CART: Variable importance for D_a

vidually can better contribute to predicting BIC changes. $Case_1$, which employed only the "Number of Customers" variable, exhibited a low predictive contribution when it was used as the only explanatory variable, but this increased when it was combined with other variables.

Second, we discuss the differences compared to the results derived by applying random forest. Combinations encompassing more than three variables were not ranked, indicating that the rules could be adequately generated using one or two variables. Because the decision tree process of random forest is more robust than that of CART, it is characterized by the extensive use of variables in discriminant rule generation.

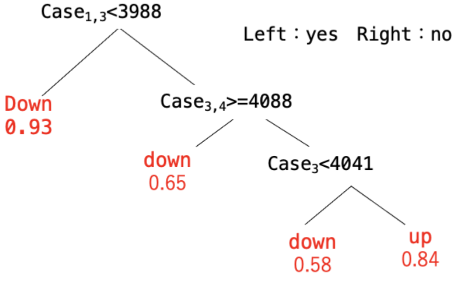


Figure 4. CART: Acquired rules for D_a

Similar to the results obtained from the random forest algorithm, $Case_{1,3}$, which was a combination of "Number of Customers" and "Number of Views on Gourmet Sites", determined the "down" trend in BIC for total sales at the highest branching point. Similarly, Tables 8 and 10 display the Gini coefficients and accuracy of the decision rules for the case of D_b, respectively, sorted in descending order.

Fourteen variable combinations were identified as contributing factors for implementing the random forest algorithm. As presented in Table 8, the variables

Combination of Variable	MeanDecreaseGini
$Case_4$	1.9333199
$Case_{2,4}$	1.5876963
$Case_2$	1.5788962
$Case_{3,4}$	1.3392425
$Case_1$	1.3273229
$Case_{1,2,4}$	1.2881951
$Case_{1,2,3}$	1.2374755
$Case_3$	1.2294474
$Case_{1,3,4}$	1.1410541
$Case_{1,4}$	1.0840145
$Case_{1,2}$	1.0289199
$Case_{2,3}$	0.9279831
$Case_{1,3}$	0.8489635
$Case_{2,3,4}$	0.7583879

Table 9. Random forest: Gini coefficient of each variable for D_b

	Down	Up	class.error
Down	14	8	0.3636364
Up	10	5	0.6666667
Total	24	13	0.4865

Table 10. Random forest: Discriminant results of rules for D_b

with high Gini coefficient values for predicting BIC changes included "Day of the Week", "Probability of Precipitation" + "Day of the Week", and "Probability of Precipitation." Similar to the case of D_a, $Case_4$, which employed only "Day of the Week", and $Case_1$, which employed only "Number of Customers", provided better predictive contributions when they were used independently rather than in combination with other variables. Furthermore, as indicated in Table 10, the error rate for distinguishing "down" trends was relatively low, yet that for distinguishing "up" trends exceeded 60 %, and the accuracy of the entire rule set was more than 50 %. Figure 5 shows the list of decision rules acquired for the total sales change.

Analogous to the case of D_a, the primary branching point, encompassing $Case_{1,3}$, "Number of Customers", and "Number of Views on Gourmet Sites", established the 'down' trend of BIC for total sales. In the prediction rules, $Case_1$, $Case_2$, $Case_{1,3}$, $Case_{3,4}$, and $Case_{1,2,4}$ primarily comprised one or two variables apart from $Case_{1,2,4}$. Finally, Table 11 lists the variable importance, sorted in descending order, and Figure 6 illustrates the discriminant rules for D_b, derived by applying the CART algorithm.

By applying the CART algorithm, eight variable combinations were identified as contributing variables. First, we discuss the similarities with the results of random forest. "Day of the Week", included in $Case_4$, was the most consequential variable

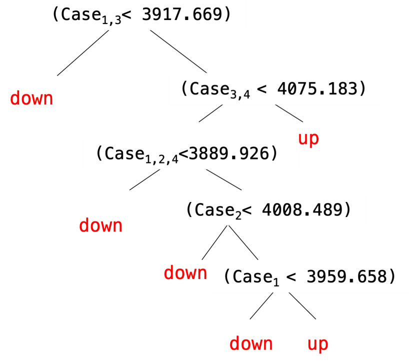


Figure 5. Random forest: Acquired rules for D_b

Variable Combination	Values
$Case_4$	20
$Case_1$	16
$Case_{1,4}$	14
$Case_{1,3}$	14
$Case_{3,4}$	14
$Case_3$	12
$Case_2$	5
$Case_{1,2}$	5

Table 11. CART: Variable importance for D_b

and its importance diminished when integrating it with other variables. Hence, employing it individually can significantly contribute predicting BIC changes. Second, we address the disparities between the results of random forest. Analogous to D_a, combinations comprising more than three elements were not ranked, indicating that discriminant rules can be adequately generated using only one or two variables.

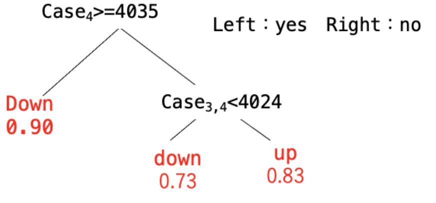


Figure 6. CART: Acquired rules for D_b

In contrast to the results of the data and methods presented thus far, the initial branching point, $Case_4$, "Day of the Week", determines the decline in BIC for total sales. From the experimental results obtained using two real datasets, we deduced that the seasonal variable should be preferably used independently and the "Probability of Precipitation" variable contributes significantly even when employed alone. Notably, although the prediction accuracy of the entire rule surpassed 50 %, further improvements can be made by incorporating enough training data.

Moreover, the outcomes obtained using the two decision tree-based algorithms indicate that CART produces discriminant rules specialized for the training data, allowing it to establish decision trees with only one or two variables. In contrast, random forest randomly partitions the training data and generates discriminant rules, thereby obtained robust decision trees with a more comprehensive range of variables. The explanatory variable, which amalgamates all variables, has not been used for decision tree generation in either the CART or random forest algorithms. This is can be attributed to the higher contribution of seasonal variables when used individually for rule generation, resulting in lower contribution when combined with other variables.

In conclusion, the accuracy of BIC prediction in multivariate time-series analyses can be enhanced by using seasonal variables individually and appropriately amalgamating other variables. The classification tree revealed the characteristic variables and their impacts on other variables; these insights that were not derived using the DTW method.

5 CONCLUSIONS

This study aimed to improve the prediction accuracy of multivariate time-series models by analyzing BIC changes. First, we evaluate the impact of variable combinations on BIC accuracy using descriptive statistics, and we discovered that the improvement conditions for the BIC of the prediction model varied depending on the combination of explanatory variables, their features, and learning periods. Moreover, we demonstrated that forecasting the estimation accuracy of a model enables more effective variable selection for model generation. Second, we identified variables with significant attributes by employing DTW to cluster and visualize the trends in BIC changes for independently used variables and their combinations.

The study results suggest that using random forest to predict the accuracy of BIC fluctuations for the dynamic regression model of ARIMA provides invaluable information to generate variable combinations. However, there is still room to enhance the discrimination accuracy for the acquired rules. Furthermore, we demonstrated the possibility of creating models with optimal accuracy by selecting variable combinations after pre-verifying their features.

In future studies, we will aim to reevaluate our findings using learning or validation data and investigate the features of the approach by comparing its performance with that of deep learning methods such as LSTM.

Acknowledgements

This work was supported by a grants-in-aid of the Uruma Fund for the Promotion of Science.

REFERENCES

[1] Box, G. E. P.—Jenkins, G. M.: Time Series Analysis: Forecasting and Control. Holden-Day, 1970.

[2] Franses, P. H.: Periodicity and Stochastic Trends in Economic Time Series. Oxford University Press, 1996, doi: 10.1093/oso/9780198774532.001.0001.

[3] Siami-Namini, S.—Tavakoli, N.—Siami Namin, A.: A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1394–1401, doi: 10.1109/ICMLA.2018.00227.

[4] Findley, D. F.—Monsell, B. C.—Bell, W. R.—Otto, M. C.—Chen, B. C.: New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. Journal of Business and Economic Statistics, Vol. 16, 1998, No. 2, pp. 127–152, doi: 10.1080/07350015.1998.10524743.

[5] Makridakis, S.—Spiliotis, E.—Assimakopoulos, V.: Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward. PLoS ONE, Vol. 13, 2018, No. 3, Art. No. e0194889, doi: 10.1371/journal.pone.0194889.

[6] Imrul Hasan, A. H. M.—Haddawy, P.: Integrating ARIMA and Spatiotemporal Bayesian Networks for High Resolution Malaria Prediction. In: Kaminka, G. A., Fox, M., Bouquet, P., Hüllermeier, E., Dignum, V., Dignum, F., van Harmelen, F. (Eds.): 22nd European Conference on Artificial Intelligence (ECAI 2016). IOS Press, Frontiers in Artificial Intelligence and Applications, Vol. 285, 2016, pp. 1783–1790, doi: 10.3233/978-1-61499-672-9-1783.

[7] Ghazvinian, H.—Mousavi, S. F.—Karami, H.—Farzin, S.—Ehteram, M.—Hossain, M. S.—Fai, C. M.—Hashim, H. B.—Singh, V. P.—Ros, F. C.—Ahmed, A. N.—Afan, H. A.—Lai, S. H.—El-Shafie, A.: Integrated Support Vector Regression and an Improved Particle Swarm Optimization-Based Model for Solar Radiation Prediction. PLoS ONE, Vol. 14, 2019, No. 5, Art. No. e0217634, doi: 10.1371/journal.pone.0217634.

[8] Wichard, J. D.: Forecasting the NN5 Time Series with Hybrid Models. International Journal of Forecasting, Vol. 27, 2011, No. 3, pp. 700–707, doi: 10.1016/j.ijforecast.2010.02.011.

[9] Khashei, M.—Bijari, M.—Ardali, G. A. R.: Hybridization of Autoregressive Integrated Moving Average (ARIMA) with Probabilistic Neural Networks (PNNs). Computers & Industrial Engineering, Vol. 63, 2012, No. 1, pp. 37–45, doi: 10.1016/j.cie.2012.01.017.

[10] Iha, T.—Kawamitsu, I.—Ohshiro, A.—Nakamura, M.: An LSTM-Based Multivariate Time Series Predicting Method for Number of Restaurant Customers in Tourism Resorts. 2021 36th International Technical Conference on Cir-

cuits/Systems, Computers and Communications (ITC-CSCC), IEEE, 2021, pp. 1–4, doi: 10.1109/ITC-CSCC52171.2021.9501432.

[11] Goya, I.—Iha, T.—Ohshiro, A.—Nakamura, M.: Predicting the Number of Customers in Resort Areas Based on Event Impact Time Series Generation. 126$^{\text{th}}$ Knowledge-Based System Study Group (SIG-KBS-126), 2022, pp. 1–5, doi: 10.11517/jsaikbs.126.0_01 (in Japanese).

[12] Zhang, Y. M.—Wang, H.—Bai, Y.—Mao, J. X.—Xu, Y. C.: Bayesian Dynamic Regression for Reconstructing Missing Data in Structural Health Monitoring. Structural Health Monitoring, Vol. 21, 2022, No. 5, pp. 2097–2115, doi: 10.1177/14759217211053779.

[13] Teixeira, J. P.—Fernandes, P. O.: Tourism Time Series Forecast – Different ANN Architectures with Time Index Input. Procedia Technology, Vol. 5, 2012, pp. 445–454, doi: 10.1016/j.protcy.2012.09.049.

[14] Claveria, O.—Torra, S.: Forecasting Tourism Demand to Catalonia: Neural Networks vs. Time Series Models. Vol. 36, 2014, pp. 220–228, doi: 10.1016/j.econmod.2013.09.024.

[15] Yuyama, A.—Hase, T.—Mizuno, T.—Maeda, S.—Takahashi, Masaya nad Tanaka, K.—Hoshino, Y.: Analysis of Prediction Error in Wind Speed Prediction Based on VAR. Proceedings of the JSPE Annual Conference/2016 JSPE Spring Conference, The Japan Society for Precision Engineering, 2016, pp. 579–580, doi: 10.11522/pscjspe.2016S.0_579 (in Japanese).

[16] Promprou, S.—Jaroensutasinee, M.—Jaroensutasinee, K.: Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand Using ARIMA Models. Dengue Bulletin, Vol. 30, 2006, pp. 99–106.

[17] Earnest, A.—Chen, M. I.—Ng, D.—Sin, L. Y.: Using Autoregressive Integrated Moving Average (ARIMA) Models to Predict and Monitor the Number of Beds Occupied During a SARS Outbreak in a Tertiary Hospital in Singapore. BMC Health Services Research, Vol. 5, 2005, No. 1, Art. No. 36, doi: 10.1186/1472-6963-5-36.

[18] Murat, M.—Malinowska, I.—Gos, M.—Krzyszczak, J.: Forecasting Daily Meteorological Time Series Using ARIMA and Regression Models. International Agrophysics, Vol. 32, 2018, No. 2, pp. 253–264, doi: 10.1515/intag-2017-0007.

[19] Díaz-Robles, L. A.—Ortega, J. C.—Fu, J. S.—Reed, G. D.—Chow, J. C.—Watson, J. G.—Moncada-Herrera, J. A.: A Hybrid ARIMA and Artificial Neural Networks Model to Forecast Particulate Matter in Urban Areas: The Case of Temuco, Chile. Atmospheric Environment, Vol. 42, 2008, No. 35, pp. 8331–8340, doi: 10.1016/j.atmosenv.2008.07.020.

[20] Nakagawa, K.—Imamura, M.—Yoshida, K.: Stock Price Prediction with Fluctuation Patterns Using Indexing Dynamic Time Warping and k^*-Nearest Neighbors. In: Arai, S., Kojima, K., Mineshima, K., Bekki, D., Satoh, K., Ohta, Y. (Eds.): New Frontiers in Artificial Intelligence (JSAI-isAI 2017). Springer, Cham, Lecture Notes in Computer Science, Vol. 10838, 2018, pp. 97–111, doi: 10.1007/978-3-319-93794-6_7.

[21] Pandey, D.—Singh, K. K.: Implementation of DTW Algorithm for Voice Recognition Using VHDL. 2017 International Conference on Inventive Systems and Control (ICISC), 2017, pp. 1–4, doi: 10.1109/ICISC.2017.8068638.

[22] Maaliw, R. R.—Quing, K. A. C.—Susa, J. A. B.—Marqueses, J. F. S.—Lagman, A. C.—Adao, R. T.—Fernando-Raguro, M. C.—Canlas, R. B.: Clustering and Classification Models for Student's Grit Detection in E-Learning. 2022 IEEE World AI IoT Congress (AIIoT), 2022, pp. 39–45, doi: 10.1109/AIIoT54504.2022.9817177.

[23] Izakian, H.—Pedrycz, W.—Jamal, I.: Fuzzy Clustering of Time Series Data Using Dynamic Time Warping Distance. Vol. 39, 2015, pp. 235–244, doi: 10.1016/j.engappai.2014.12.015.

[24] Akaike, H.: Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B. N., Caski, F. (Eds.): Proceeding of the Second International Symposium on Information Theory. Akademiai Kiado, Budapest, 1973, pp. 267–281.

[25] Schwarz, G.: Estimating the Dimension of a Model. The Annals of Statistics, Vol. 6, 1978, No. 2, pp. 461–464, doi: 10.1214/aos/1176344136.

[26] Albahli, S.—Yar, G. N. A. H.: Defect Prediction Using Akaike and Bayesian Information Criterion. Computer Systems Science and Engineering, Vol. 41, 2022, No. 3, pp. 1117–1127, doi: 10.32604/csse.2022.021750.

[27] Kasali, J.—Adeyemi, A. A.: Model-Data Fit Using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the Sample-Size-Adjusted BIC. Square: Journal of Mathematics and Mathematics Education, Vol. 4, 2022, No. 1, pp. 43–51, doi: 10.21580/square.2022.4.1.11297.

[28] Hyndman, R. J.—Athanasopoulos, G.: Forecasting: Principles and Practice. 3rd Edition. OTexts, 2021, `https://otexts.com/fpp3/`.

[29] Breiman, L.: Random Forests. Machine Learning, Vol. 45, 2001, No. 1, pp. 5–32, doi: 10.1023/A:1010933404324.

[30] Breiman, L.—Friedman, J.—Stone, C. J.—Olshen, R. A.: Classification and Regression Trees. Taylor & Francis, 1984.

Ayako OHSHIRO received her B.Sc. and M.Sc. degrees from the University of the Ryukyus in 2009 and 2011, respectively. She took her Ph.D. from the University of Ryukyus in 2017 and she has been Associate Professor at the Okinawa International University. Her research interests include data analysis at the field of intelligent information engineering.

Morikazu NAKAMURA received his B.E. and M.E. degrees from the the University of the Ryukyus in 1989 and 1991, respectively, and his D.E. degree from the Osaka University in 1996. He is currently Professor of computer science and intelligent systems at the Faculty of Engineering, University of the Ryukyus, Japan. His research interests include the theory and applications of mathematical systems. He is a member of IEICE and IEEE.