Journal of Engineering Science and Technology
Vol. 19, No. 2 (2024) 390 - 404

BEST CLASSIFIER OVER FEATURE SELECTION AND DELTA ERROR IN TIME SERIES DATASET

NUGROHO WIDYANTO[1,*], JIN-WHAN KIM[1],
AGUS NURSIKUWAGUS[2]

[1]Department of Computer Engineering, Youngsan University, 288 Junam-ro, Yangsan, Gyeongsangnam-do, South Korea
[2]Department of Information System, Universitas Komputer Indonesia, Indonesia
*Corresponding Author: nugrohowidyanto@office.ysu.ac.kr

Abstract

The selection of the best classifier on conventional machine learning is often made intuitively by observing existing research. It becomes an obstacle when stating high accuracy, regardless of whether the machine suits the dataset. This research proposes a method to deal with the accuracy of using a learning machine so that the accuracy generated can be aligned with the usage of the dataset. This method uses the feature selection method combined with the machine learning classifier. SelectKbest and principal component analysis (PCA) methods combined with the classifier machine by counting confirmed MSE, MAE, R^2, and delta errors can predict the appropriate machine learning for the OTHERS datasets. Delta error and R^2 by SelectKBest and the PCA can see the Delta Error tendency of its classifier learning. The observed classifiers confirmed the best R^2 values on linear and Bayesian regression. The R^2 values of Bayesian and linear regression are 0.634 and 0.687, respectively. Average Delta Error of MAE of SelectiKbest < Average Delta Eror of MAE of PCA, $3.21\times10^{15} < 1.37\times10^{16}$. In future research, we found the challenge of the best classifier in Deep Learning and the method of selecting features appropriate for the time series dataset.

Keywords: Chi-square, Classifier, Delta error, Principal component analysis, SelectKBest.

1. Introduction

COVID-19 has become one of the most dangerous diseases that has a widespread impact in the world. Millions of people have died from this disease. The movement of suffering daily continues to increase in various countries. The disease also has a significant impact on people with or without comorbidities such as diabetes, heart, hypertension, and other comorbid diseases. The progression of this disease is recorded on the Kaggle dataset OTHERS or owid-data-covid19 [1-3]. A time series dataset is used to predict an increase in Covid-19 cases. Some research proposals can be proposed in supervising machine learning with data labelled and time series. OTHERS multivariate or multilabel data values can be studied from various sides.

The problem interesting in the research is how to determine the appropriate classifier for OTHERS datasets by parameters obtained. The parameters have been proposed, such as MSE, MAE, $R^{2,}$ and the Delta Error, which are extracted by feature reduction like Principal Component Analysis (PCA) [1] and SelectKBest [2]. The use of classifiers for a data set needs to be considered, such as data values, data collection techniques whether time series or cross-sectional, encoding methods, Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and chi-square (R^2) [2-7]. The considerations presented in the paragraph above are to be a support for the solution. Leveraging methods such as encoding, MAE, MSE, RMSE, and R^2 follow the state of the art regarding the use of classifier machines. The use of PCA (1)and SelectKBest [2] for feature reduction is one of the research projects proposed to obtain the best classifier [1].

Regarding the problems and solutions proposed, the contribution of this research can be written as follows:

i) PCA and SelectKBest as a feature reduction can minimize the values of MAE, MSE, RMSE and R^2 [1]. PCA can improve the time effectiveness of obtaining the correct classifier for OTHERS datasets with the ability to minimize features according to the user's request. Similarly, with the SelectKBest method, the selection of the regression score function helps to select the right feature for the classifier used.

ii) Classifiers based on linear functions are targeted as classifiers that can predict on OTHERS datasets. The influence of Linear Regression, Support Vector Machine (SVM) [2, 8, 9], Sigmoid Gradient Descent (SGD) Regressor, and Bayesian Regression functions [2, 9-13] on predictions by relying on the PCA and SelectKBest reduction methods against OTHERS data sets, have used as an experiment to determine which classifier is the most accurate.

iii) The calculation of the minimum Delta Error has been used as a restriction to determine the best classifier for the OTHERS datasets. Delta error is a combination of errors generated by the PCA classifier [1] and SelectKBest methods. Delta Errors such as delta error for MSE, MAE, RMSE, and R^2 were targeted.

Presentations and writings in this article are structured based on the systematics of writing so the readers can easily understand. The first section is an introduction that contains a write of the problems and contributions. The second section is a related work that contains state-of-the-art. The third section is methods, which explains the proposal of steps. The fourth section is the outcome and discussion, which presents the results and discussion. The fifth section is a conclusion that is a summary of the entire presentation in this article.

2. Related Works

Machine learning is divided into several types of learning depending on the data it receives. Machine learning can be subdivided into supervised, Unsupervised, Reinforcement, and Association [9, 14]. A Machine classifier works on data sets labelled or has classes. Variations of this classifier machine are also divided again into prediction and regression.

Predictions are often associated with discrete data types, whereas regression is associated with numeric data types with time series data. In connection with this research, the machine learning used is a regression machine because the data used is a time series type. The regression function predicts the likelihood of coming with the numeric data it holds. Some other variations of regression functions used are Bayesian regression, SVM, Linear Regression, and SGD regressor [8].

3. Research Methods

This section presents the methods used in this research. This explanation covers the model pipeline of the research and the supporting methods.

The use of methods to solve classification issues is a step that must be demonstrated to know how research works. This model pipeline is a guide to the steps of work when solving a question. Figure 1 depicts the research method for accomplishing the experiments.

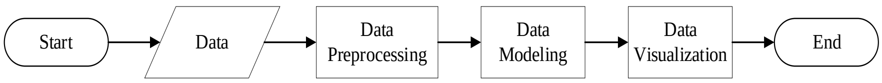

Fig. 1. Research methods.

3.1. Data preprocessing

Data preprocessing is part of the process, including data cleaning, missing values, and Exploratory Data Analysis (EDA) [15]. EDA is a method of data analysis that emphasizes the patterns of data sets used. EDA processes include descriptive statistics, correlation [16], and feature selection. Some statistical methods are also used to determine the validity of a data set. The formula used is to calculate the average, maximum value, minimum value, standard deviation, and correlation [16].

$$mean(x) = \frac{\sum_i^n x_i}{N} \quad (1)$$

$$Max = Max[x_1 \dots x_n] \quad (2)$$

$$Min = Min[x_1 \dots x_n] \quad (3)$$

$$S = \frac{\sqrt{\sum f_i (x_i - \bar{x})^2}}{\sum f_i} \quad (4)$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (5)$$

where r = Pearson correlation coefficient, x_i = x variable samples, $\bar{x}$ = mean of values in x variable, y_i = y variable samples, $\bar{y}$ = mean of values in y variable.

Correlation can describe the impact of a large value, so it can be stated that according to Evans (1996) gives dependence between variables with a vulnerability of 0.00-0.19 "very weak", 0.20-0.39 "weak", 0.40-0.59 "moderate", 0.60-0.79 "strong", 0.80-1.0 "very strong" [16].

Other metrics to encourage the classifier are Mean Square Error (MSE), Mean Average Error (MAE), Root Mean Square Error (RMSE), and chi-square (R^2). The formulation can be written as Eqs. (6) to (9)

$$MSE = \frac{\sum(y_i - \bar{y}_i)^2}{n} \tag{6}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum(y_i - \bar{y}_i)^2}{n}} \tag{8}$$

$$R^2 = \sum \frac{(O-E)^2}{E} \tag{9}$$

where: R^2 = chi-square test statistics, Σ = cumulative operators (meaning “take the number”); O = observed frequency; E = expected frequencies.

3.2. Rotation

The second method is rotation or rotating the image clockwise or turning clockwise by using Eq. (10).

$$\begin{aligned} x1 = integer(\cos\cos\theta * \left(x2 - \left(\frac{w}{2}\right)\right) - \sin\sin\theta * \left(y2 - \left(\frac{h}{2}\right)\right) + \left(\frac{w}{2}\right)\ )y1 \\ = \ \theta * \left(x2 - \left(\frac{w}{2}\right)\right) + \cos\cos\theta\ * \left(y2 - \left(\frac{h}{2}\right)\right) + \left(\frac{h}{2}\right)) \end{aligned} \tag{10}$$

3.3. Data modelling

The sub-section describes using the classifier machine as a prediction machine on the OTHERS datasets. The proposed classifier is Linear Regression, Support Vector Machine (SVM), Stochastic Gradient Descent Regressor (SGD Regressor), and Bayesian Regressors [8, 10, 11, 13]. The working model of this learning machine can be seen in Fig. 2.

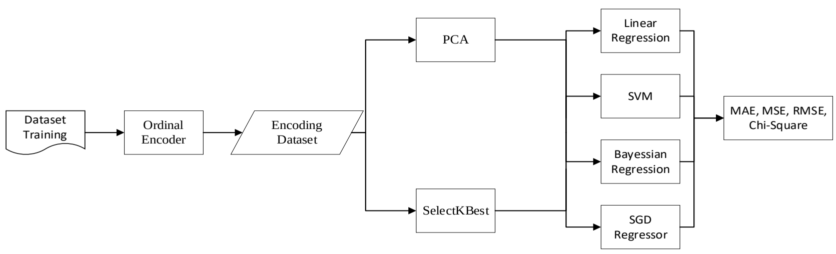

Fig. 2. Classifier model.

Linear regression is one of the most important and widely used regression techniques. It is one of the simplest regression methods. Regression is about determining the best prediction weight that matches the smallest residue. To obtain the best weight, usually minimize the number of square residues (SSR) for all observations i = 1,...; n: SSR = $\Sigma i(y_i - f(x_i))^2$ [11]. This approach is called the common smallest square method [11, 12].

This shape is optimized directly by Linear Support Vector Classification (SVC) or Sigmoid Gradient Descent (SGD) Classifier. Still, unlike a double shape, this shape does not involve a product in between samples, so the famous kernel tricks cannot be applied. This is why only linear kernels are supported by Linear SVC (identity function) and can be written as below [8]:

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{n} max(0,1 - y_i(w^T \phi(x_i) + b)) \quad (11)$$

Principal Component Analysis (PCA) [1] and SelectKBest as a feature/ variable reduction method are used to obtain predictive results effectively. The number of reduced components starts from two to nineteen. PCA and SelectKbest reduction results are then calculated hybrid error using the following proposed formulation [1].

$$\overline{Error}_{MSE} = 1 - \left[\frac{(\overline{MSE}_{SelectKBest} * \overline{MSE}_{PCA})}{(\overline{MSE}_{SelectKBest} + \overline{MSE}_{PCA})}\right] \quad (12)$$

$$\overline{Error}_{MAE} = 1 - \left[\frac{(\overline{MAE}_{SelectKBest} * \overline{MAE}_{PCA})}{(\overline{MAE}_{SelectKBest} + \overline{MAE}_{PCA})}\right] \quad (13)$$

$$\overline{Error}_{RMSE} = 1 - \left[\frac{(\overline{RMSE}_{SelectKBest} * \overline{RMSE}_{PCA})}{(\overline{RMSE}_{SelectKBest} + \overline{RMSE}_{PCA})}\right] \quad (14)$$

If we want to know if the error is close to the linearity, then the delta error is calculated for each change value. To calculate the delta error on each PCA variable selection method or SelectKBest, we can use the formula proposed as follows [4, 9, 17]:

$$\Delta MSE_{SelecKBest_i} = ABS\left(Error_{MSE} - MSE_{SelectKBest_i}\right) \quad (15$$

$$\Delta MAE_{SelecKBest_i} = ABS\left(Error_{MAE} - MAE_{SelectKBest_i}\right) \quad (16$$

$$\Delta RMSE_{SelecKBest_i} = ABS\left(Error_{RMSE} - RMSE_{SelectKBest_i}\right) \quad (17)$$

$$\Delta MSE_{PCA_i} = ABS\left(Error_{MSE} - MSE_{PCA_i}\right) \quad (18)$$

$$\Delta MAE_{PCA_i} = ABS\left(Error_{MAE} - MAE_{PCA_i}\right) \quad (19$$

$$\Delta RMSE_{PCA_i} = ABS\left(Error_{RMSE} - RMSE_{PCA_i}\right) \quad (20)$$

4. Results and Discussion

This section describes the results of experiments with the proposed model. Exhibitions related to data sets include Exploratory Data Analysis (EDA), modelling, and visualization.

4.1. Dataset

This dataset is taken from the COVID-19 survey. This dataset has 134,020 instances with 19 attributes/variables as independent and one as a class target. The dataset used is taken from worldwide COVID data collected in the years 2020 – 2021 in a time series. Dataset properties can be seen in Table 1 [8].

Table 1 is used as a data set in experiments using machine learning classifiers. Table 1 is also accompanied by a class/label named "total cases" that records the evolution of the number of cases from day to day. The data set is divided into 70:30, 93814 training and 40206 testing datasets.

Table 1. Dataset other properties.

Attribute	Describes
Iso_code	ISO 3166-1 alpha-3 – three-letter country codes. OWID-defined regions (e.g., continents like 'Europe') contain the prefix 'OWID_'.
Continent	Continent of the geographical location
Location	Geographical location. Location 'International' considers special regions ("Diamond Princess" and "MS Zaandam" cruises).
Date	Date of observation
Population	Population (latest available values). See https://github.com/owid/covid-19-data/blob/master/scripts/input/un/population_latest.csv for the full list of sources.
Population_Density	Number of people divided by land area, measured in square kilometres, most recent year available
Median_age	Median age of the population, UN projection 2020
Age_65_older	Share of the population that is 65 years and older, most recent year available
Age_70_older	Share of the population that is 70 years and older in 2015
Gdp_per_capita	Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available
Extreme_proverty	Share of the population living in extreme poverty, most recent year available since 2010
Cardiovasc_death_rate	Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people)
Diabetes_prevelence	Diabetes prevalence (% of population aged 20 to 79) in 2017
Female_smokers	Share of women who smoke, most recent year available
Male_smokers	Share of men who smoke, most recent year available
Handwashing_facilities	Share of the population with basic handwashing facilities on premises
hospital_beds_per_thousand	Hospital beds per 1,000 people, most recent year available since 2010
life_expectancy	Life expectancy at birth in 2019
human_development_index	A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living. Values for 2019, imported from http://hdr.undp.org/en/indicators/137506

4.2. Exploratory data analysis (EDA)

Some of the analyses applied were to check the attribute value on the dataset for null value, data type check, and the measurement of statistical values such as averages, standard deviation values, and quarters. The model pipeline for this EDA [15] can be seen in Fig. 3. Table 2 results from Null checks on each attribute in the OTHERS dataset.

The next improvement in the null/missing value process is to give the missing value a specific value. Table 3 is the result of verifying OTHERS datasets that are already free of missing values.

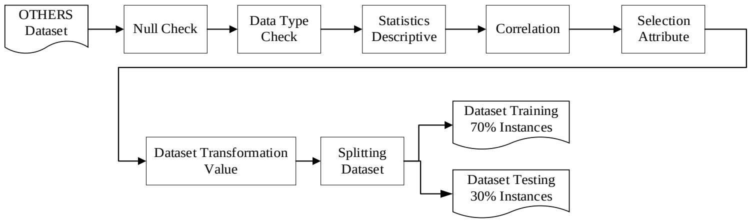

Fig. 3. Pipeline EDA.

Table 2. Name variable has a null value (NaN).

Variable	Total NaN value
excess_mortality_cumulative_absolute	5031
excess_mortality_cumulative	5031
excess_mortality	5031
excess_mortality_cumulative_per_million	5031

Table 3. Name attribute has a clean from null (NaN).

Variable	Total NaN value
excess_mortality_cumulative_absolute	5031
excess_mortality_cumulative	5031
excess_mortality	5031
excess_mortality_cumulative_per_million	5031

Table 4 verifies the data type of the value of the OTHERS table. The OTHERS dataset value consists of the object, the discrete data value, the float64 data value, and a numerical data value. The next process is transforming the multitype data value into a homogeneous data value to be more efficient in processing. This transformation process uses the ordinal encoder method. Ordinal Encoder is a method of transforming data values from categorical to numeric. Table 5 shows data types after transformation.

Table 4. Attributes data type.

Attribute	Type
iso_code	object
continent	object
date	object
…	…
excess_mortality	float64
excess_mortality_cumulative_per_million	float64

Table 5. Transformation results by ordinal encoding [9].

Variable	Types	Variable	Types
total_cases	float64	gdp_per_capita	float64
total_vaccinations	float64	extreme_poverty	float64
continent	float64	cardiovasc_death_rate	float64
location	float64	diabetes_prevalence	float64
date	float64	female_smokers	float64
population	float64	male_smokers	float64
population_density	float64	handwashing_facilities	float64
median_age	float64	hospital_beds_per_thousand	float64
aged_65_older	float64	life_expectancy	float64
aged_70_older	float64	human_development_index	float64

Table 6 is a statistical description of the OTHERS dataset. Descriptive statistics give a sum, average, standard deviation, minimum, and maximum value.

Table 6. Statistics descriptive [15].

Attributes	**count**	**mean**	**std**	**min**	**max**
total_cases	1.34×10^5	1.96×10^{12}	1.15×10^{13}	10^6	2.55×10^{14}
new_cases	1.34×10^5	7.93×10^9	4.25×10^{10}	-7.45×10^7	9.08×10^{11}
new_cases_smoothed	1.34×105	7.88×10^9	4.19×10^{10}	-6.22×10^6	8.27×10^{11}
total_deaths	1.34×105	4.40×10^{10}	2.44×10^{11}	106	$5.12\times10_{12}$
...	...	...	...	...	...
excess_mortality	1.34×105	5.44×10^6	2.56×10^7	-9.6×10^7	3.74×10^8
excess_mortality_cumulative_per_million	1.34×105	2.41×10^8	8.72×10^8	-1.73×10^{10}	6.17×10^6

Regarding EDA [15], the study counts the correlation between each variable. Figure 4 shows the Pearson correlation between the variables from the OTHERS dataset. Table 7 is the result SelectKBest for K=2, ...10.

Table 8 provides an overview of attributes similar to the SelectKbest and Pearson Correlation process results. The attribute selection test in Table 7 only determines which attributes are the most dominant in determining the class "total_cases". In the Pearson correlation process obtained, nine variables have a significant dimension > 0.05.

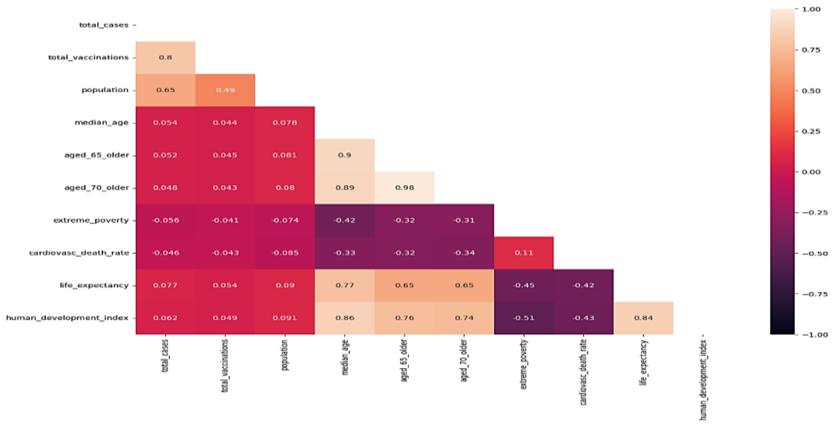

Fig. 4. Nine attributes are significantly higher than α = 0,05 [16].

Table 7. Attributes selected by SelecKBest.

K	**Attribute Selected**	**Score Function**
1	total_vaccinations, population	F_regression
2	total_vaccinations, date, population	F_regression
3	total_vaccinations, date, population, cardiovasc_death_rate	F_regression
4	total_vaccinations, date, population, aged_65_older, cardiovasc_death_rate	F_regression
5	total_vaccinations, date, population, aged_65_older, aged_70_older, cardiovasc_death_rate	F_regression
6	total_vaccinations, continent, date, population, aged_65_older, aged_70_older, cardiovasc_death_rate	F_regression
7	total_vaccinations, continent, date, population, aged_65_older, aged_70_older, cardiovasc_death_rate, female_smokers	F_regression
8	total_vaccinations, continent, date, population, aged_65_older, aged_70_older, cardiovasc_death_rate, female_smokers, hospital_beds_per_thousand	F_regression
9	total_vaccinations, continent, location, date, population, aged_65_older, aged_70_older cardiovasc_death_rate, female_smokers, hospital_beds_per_thousand	F_regression

Table 8. Comparison selection attributes.

Attributes (Over correlation methods)	Attribute (over SeletedKBest methods K= 10)
total_cases	total_cases
total_vaccinations	total_vaccinations
population	population
median_age	date
aged_65_older	aged_65_older
aged_70_older	aged_70_older
extreme_poverty	continent
cardiovasc_death_rate	cardiovasc_death_rate
life_expectancy	female_smokers
human_development_index	hospital_beds_per_thousand

The equation formed by 20 variables can be written in Table 8 [11, 12] based on linear regression calculations. This study is noted by one variable intercept and 19 independent variables. The variables are an intercept, total_vaccinations (X1), continent (X2), location X3, date X4, population (X5), population_density(X6), median_age (X7), aged_65_older (X8), aged_70_older (X9), gdp_per_capita (X10), extreme_poverty (X11), cardiovasc_death_rate (X12), diabetes_prevalence (X13), female_smokers (X14), male_smokers (X15), handwashing_facilities (X16), hospital_beds_per_thousand (X17), life_expectancy (X18), human_development_index (X19) [11, 12].

$$\begin{aligned} f(x) = {} & -39609{,}12 + 0{,}66X_1 + 1432{,}84\,X_2 - 9{,}72X_3 + 37{,}46X_4 + 201{,}58X_5 \\ & + 14{,}45X_6 - 17{,}05X_7 + 11{,}11X_8 - 4{,}21X_9 + 11{,}81X_{10} \\ & - 6{,}80X_{11} + 23{,}18X_{12} + 5{,}55X_{13} + 15{,}44\,X_{14} + 22{,}00X_{15} \\ & - 7{,}53X_{16} - 34{,}20X_{17} - 10{,}56X_{18} + 162{,}82X_{19} \end{aligned} \tag{21}$$

4.3. Results

This section shows the results of experiments using the Linear Regression classifier machine. The experimental scenarios are as follows:

First scenario:

i). Use the SeleckKBest method with a score function regression starting with K= 2 ,...; 10, 20.

ii). Do predictions using a Linear Regression Classifier with a linear function

iii). Count MSE, MAE, and RMSE for the SelectKBest Method with Linearregression Klassifier.

Second scenario:

i). Use the Principal Components Analysis (PCA) method starting with components = 2,..;10,20.

ii). Make predictions using a Linear Regression Classifier with a straight-line function.

iii). Calculate MSE, MAE, and RMSE for PCA methods with Linear Regression Classifiers [4-6, 17, 18].

In Table 9, it is seen that there is a comparison result by SelectKBest and PCA. Both techniques are still used in classifier modelling to see which attributes have MAE, MSE, and RMSE.

Table 9. Comparison MSE, MAE, and RMSE.

K	MSE	MAE	RMSE	Classifier
2	2.77×10^{-5}	1.46×10^{5}	1.67×10^{5}	LinearRegression
3	1.51×10^{-4}	1.07×10^{6}	1.23×10^{6}	LinearRegression
4	2.27×10^{-4}	1.32×10^{6}	1.51×10^{6}	LinearRegression
5	2.12×10^{-4}	1.27×10^{4}	1.46×10^{6}	LinearRegression
6	2.87×10^{-5}	1.48×10^{5}	1.69×10^{6}	LinearRegression
7	2.12×10^{-4}	1.27×10^{6}	1.46×10^{6}	LinearRegression
8	1.87×10^{-4}	1.20×10^{6}	1.37×10^{6}	LinearRegression
9	1.80×10^{-4}	1.17×10^{6}	1.34×10^{6}	LinearRegression
10	2.22×10^{-4}	1.30×10^{5}	1.49×10^{6}	LinearRegression
20	1.71×10^{16}	1.02×10^{16}	1.31×10^{16}	LinearRegression

Table 10 shows other models regarding the use of the latent variable. The method used is Principal Component Analysis (PCA) and metric by Eqs. (6) to (8). Based on Table 9 and Table 10. Determining the MSE Average, MAE Averages, and RMSE Annual is possible. Table 11 shows the error averages of both Selection Feature methods, Eqs. (11) to (13).

4.4. Discussion

This section discusses the surroundings related to datasets, machine learning, attribute selection methods, and results obtained. Discourse highlights the use of methods and the results of using methods used for OTHERS datasets.

There are 19 latent or independent attributes, as shown in Table 1. Values of different types are then transformed using the ordinal encoder method. The impact of using the ordinal encoder method is that all data of discrete or numeric type is transformed into an integer value that has no relationship with one another. The advantage of using this data type for the OTHERS dataset is that all values are converted into integer values without restrictions. The disadvantage of using ordinal encoding is that if the data value varies tightly, it produces an integer value.

Table 10. MSE, MAE, and RMSE metrics for PCA [1].

PCA (Components)	MSE	MAE	RMSE	Classifier
2	2.96×1016	1.34×1016	1.72×1016	LinearRegression
3	2.90×1015	1.32×1016	1.70×1016	LinearRegression
4	2.83×1014	1.30×1016	1.68×1015	LinearRegression
5	2.73×1015	1.27×1016	1.65×1015	LinearRegression
6	2.54×1016	1.21×1016	1.59×1015	LinearRegression
7	2.19×1016	1.13×1016	1.48×1015	LinearRegression
8	1.79×1016	1.04×1016	1.34×1015	LinearRegression
9	1.78×1015	1.03×1016	1.34×1016	LinearRegression
10	1.78×1016	1.03×1016	1.33×1016	LinearRegression
20	1.71×1015	1.02×1016	1.31×1015	LinearRegression

Table 11. MSE, MAE, and RMSE average in SelectKBest and PCA using Eqs. (11) to (13).

$\overline{Error}_{MSE}$	$\overline{Error}_{MAE}$	$\overline{Error}_{RMSE}$
1.50×10^{15}q	-9.38×10^{14}	-1.10×10^{15}

Tables 9 and 10 are the MSE, RMSE, and MAE results. Measurements using MSE for SelectKBest are obtained relatively far from the linear line. Table 11 shows measurements PCA with MSE yield MSE larger than SelectkBest, so the error is far from linear. Figure 5 shows the reduction method with PCA by giving a new value, which yields an unoptimal result compared to SelectKBest.

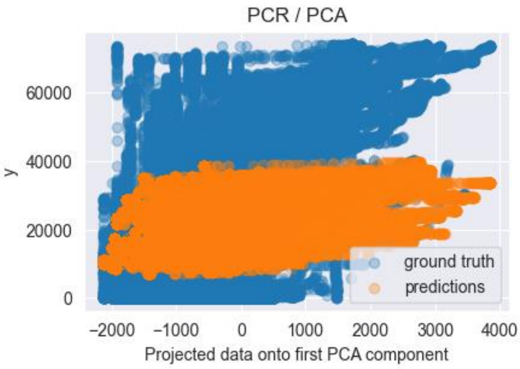

Fig. 5. Distribution PCA of linear regression classifier [1, 11, 12].

The calculation of Delta Error MSE for K=2 in SelectKBest methods using Eq. (22) is:

$$\Delta MSE_{SelecKBest_{k=2}} = ABS(-1{,}50 \times 10^{15} - 2{,}77 \times 10^{5}) = 1{,}50 \times 10^{15} \quad (22)$$

$$\Delta MAE_{SelecKBest_{k=2}} = ABS(-9{,}38 \times 10^{14} - 1{,}34 \times 10^{16}) = 9{,}38 \times 10^{14} \quad (23)$$

$$\Delta RMSE_{SelecKBest_{k=2}} = ABS(-1{,}10 \times 10^{15} - 1{,}67 \times 10^{5}) = 1{,}50 \times 10^{15} \quad (24)$$

Average Delta Error using SelectKBest:

$$\overline{\Delta MSE_{SelecKBest_{k=[1..n]}}} = \sum_i^n \frac{\Delta MSE_{SelecKBest_{k=i}}}{N} = \frac{\Delta MSE_{SelecKBest_{k=1}} + \cdots + \Delta MSE_{SelecKBest_{k=11}}}{11} = 3{,}21 \times 10^{15} \quad (25)$$

Average Delta error using PCA methods:

$$\overline{\Delta MSE_{PCA_{n=[1..n]}}} = \sum_i^n \frac{\Delta MSE_{PCA_n}}{N} = \frac{\Delta MSE_{PCA_1} + \cdots + \Delta MSE_{PCA_n}}{11} = 1{,}37 \times 10^{16} \quad (26)$$

Tables 11 and 12 show that the SelectKBest method is more stable in the error value. The error value on each K shows a homogeneous value. Average Delta Error SelectiKbest < Averages Delta Errors PCA, $3.21\times10^{15} < 1.37\times10^{16}$. This confirms that the SelectKBest method, when hybridized with the classifier machine, has predictions that are close to actual data. This argument can be confirmed by looking at delta error values on any K or feature specified by the SelectKBest method. Determining the number of components that use PCA and then looking at the predictions, the results are unstable at a certain point.

Measurements with MSE, MAE, RMSE, and R^2 show the impact of the issue of using feature selection methods. Based on the formation of functions with linear regression with various methods of regularization, the latent number of variables/features/attributes of as many as 19 with class target total_cases can be seen in Table 13. The values on the MSE and R^2 columns have the same results despite using different Regularization.

Table 12. Delta error PCA [1].

PCA (Component)	ΔMSE	ΔMAE	ΔRMSE
2	3.11×10^{16}	1.43×10^{16}	1.83×10^{16}
3	4.40×10^{15}	1.41×10^{16}	1.81×10^{16}
4	1.78×10^{15}	1.39×10^{16}	2.78×10^{15}
5	4.23×10^{15}	1.36×10^{16}	2.75×10^{15}
6	2.69×10^{16}	1.30×10^{16}	2.69×10^{15}
7	2.34×10^{16}	1.22×10^{16}	2.58×10^{15}
8	1.94×10^{16}	1.13×10^{16}	2.44×10^{15}
9	3.28×10^{15}	1.12×10^{16}	1.45×10^{16}
10	1.93×10^{16}	1.12×10^{16}	1.44×10^{16}
20	3.21×10^{15}	1.11×10^{16}	2.41×10^{15}

On other classifiers, such as the Support Vector Machine (SVM) present in Table 14, it can be noted that the values of the columns R^2 show negative values. SVM is not suitable as a classifier for the OTHERS data set case. R^2 assures us that the SVM occur overfit when used on OTHERS data sets; see Table 15. The characteristic aspect of the data set type is considered when using the classifier.

Table 13. MSE, RMSE, MAE, and R^2 using various regularization.

Model	MSE	RMSE	MAE	R^2	Regularization
Linear Regression	1.71×10^{16}	1.31×10^{15}	1.02×10^{16}	6.87×10^{-1}	No Regularization
Linear Regression	1.71×10^{16}	1.31×10^{15}	1.02×10^{16}	6.87×10^{-1}	L2
Linear Regression	1.71×10^{16}	1.31×10^{16}	1.02×10^{16}	6.87×10^{-1}	L1
	1.71×10^{16}	1.31×10^{16}	1.02×10^{15}	6.87×10^{-1}	Elastic Net Regression

Table 14. MSE, RMSE, MAE, and R^2 by SVM using various kernels.

Model	MSE	RMSE	MAE	R^2	Kernel
SVM	7.40×10^{15}	2.72×10^{15}	2.46×10^{15}	-3.50×10^{-1}	Linear
SVM	7.29×10^{15}	2.70×10^{15}	2.44×10^{16}	-3.31×10^{-1}	Polynomial
SVM	7.40×10^{15}	2.72×10^{15}	2.46×10^{15}	-3.50×10^{-1}	RBF
SVM	7.41×10^{15}	2.72×10^{16}	2.46×10^{16}	-3.51×10^{-1}	Sigmoid

Table 15. Comparison among machine learning in linear models.

Model	MSE	RMSE	MAE	R^2
SDGRegressor	1.73×10^{33}	4.16×10^{16}	3.63×10^{16}	-3.75×10^{41}
BayesianRegressor	1.99×10^{16}	1.41×10^{16}	1.08×10^{15}	6.34×10^{-1}
Linear Regression	1.71×10^{16}	1.31×10^{15}	1.02×10^{16}	6.87×10^{-1}
SVM	7.40×10^{15}	2.72×10^{15}	2.46×10^{15}	-3.50×10^{-1}

Table 15 compares selection attributes, which is another consideration in determining results. Processes in data mining become its consideration when it comes to multiple attributes. In the OTHERS dataset, 19 attributes are used as latent variables in determining the result. On the other hand, when it comes to data collection, data collected in time series is an essential consideration when using the classifier. The utilization of the best classifier and optimal feature selection can significantly contribute to enhancing the localization accuracy in an improved signal strength-based algorithm, as proposed in the preceding study [19]. The integration of these methods may reinforce the reliability of localization recognition in indoor wireless sensor networks [19].

5. Conclusion

Based on the proposed contributions, this study produces some confirmation related to the experiment being done. Machine learning suitable for OTHERS time series datasets is a regression function with a linear function. The selection of the proposed classifier machine obtains different error values. The MSE values for SDGRegressor, BayesianRegressors, Linear Regressions, and SVM are 1.73×10^{33}, 1.99×10^{16}, 1.71×10^{16} and 7.40×10^{15}, respectively. The MAE values of SDG regressors and Bayesians Regressors are 3.63×10^{16}, 1.08×10^{15}, 1.02×10^{16} and 2.46×1015, respectively. The R^2 values in SDG Regressor and Bayesian Regression are -3.75×10^{41}, 6.34×10^{-1}, 6.87×10^{-1} and -3.50×10^{-1}, respectively. Using the PCA and SelectKBest feature selection methods after performing multiple experiments with K on the SelectKBest and N on PCA components, the most stable Delta Error value is at Select KBest.

In comparison, R^2 values obtained non-negative chi-square values in the machine classifiers BayesianRegressor and Linear Regression, which are 0.634 and 0.687, respectively. The value R^2 is a measure of the fitness of a machine classifier. The experiment conducted in this study confirms that the Bayesian Regressor and Linear Regression are classifier machines suitable for other datasets, including a hybrid of Delta Error.

Using the R^2 metric helps determine the correct classifier for the dataset time series. Predictions with Linear Regression with confirmed feature selection are a suitable classifier machine for OTHERS datasets. Other outcomes that need to be discussed are data set transformation methods, multimodal or multivariate data set use, and limited feature class/target/label class. Besides, the study can also use a deep learning machine to test that. Those things can be a new challenge for research.

Nomenclatures

$\overline{Error}_{MAE}$	Hybrid Error MSE by PCA and SelectKBest
$\overline{Error}_{MSE}$	Hybrid Error MSE by PCA and SelectKBest
$\overline{Error}_{RMSE}$	Hybrid Error MSE by PCA and SelectKBest
f_i	Frequencies at i-th
Max	Maximum value from data series
$mean(x)$	Average of data series
Min	Minimum value from data series
N	Population number
R^2	Chi-Square
r	Pearson Correlation
S	Standard deviation
x_i	Single data at i-th

Greek Symbols

ΔMSE_{PCA_i}	Delta Error PCA
$\overline{\Delta MSE_{SelecKBest_{k=[1..n}}}$	Average of Delta Error for each SelectKBest
$\phi(x_i)$	Linear Function

Abbreviations

MAE	Mean Absolute Error
MSE	Mean Square Error
PCA	Principal Component Analysis
RMSE	Root Mean Square Error

References

1. Liu, Y.; Wang, Z.; Yu, X.; Chen, X.; and Sun, M. (2022). Memory-based transformer with shorter window and longer horizon for multivariate time series forecasting. *Pattern Recognition Letters*, 160(1), 26-33.
2. Ahmed, S.; Nielsen, I.E.; Tripathi, A.; Siddiqui, S.; Ramachandran, R.P.; and Rasool, G. (2023). Transformers in time-series analysis: A tutorial. *Circuits, Systems, and Signal Processing*, 42(12), 7433-7466.
3. Lee, T.L.; and Jeng, D.S. (2002). Application of artificial neural networks in tide-forecasting. *Ocean Engineering*, 29(9), 1003-1022.
4. Tipping, M.E.; and Bishop, C.M. (1999). Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2), 443-482.
5. Yue, Z.; Sun, C.; Gao, L.; Zhang, Y.; Shabbir, S.; Xu, W.; and Yu, J. (2020). Machine learning efficiently corrects LIBS spectrum variation due to change of laser fluence. *Optics Express*, 28(10), 14345-14356.
6. Liu, M.; Zeng, A.; Chen, M.; Xu, Z.; Lai, Q.; Ma, L.; and Xu, Q. (2022). Scinet: Time series modelling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35(1), 5816-5828.
7. Kusumaningrum, N.; Purnamasari, I.; and Siringoringo, M. (2023). Peramalan menggunakan model hybrid ARIMAX-NN untuk total transaksi pembayaran nontunai. *VARIANSI: Journal of Statistics and its Application on Teaching and Research*, 5(1), 1-15.
8. Parhusip, H.A. (2020). Study on COVID-19 in the world and Indonesia using regression model of SVM, Bayesian ridge and gaussian. *Jurnal Ilmiah Sains*, 20(2), 49-59.
9. Gambella, C.; Ghaddar, B.; and Naoum-Sawaya, J. (2021). Optimization problems for machine learning: A survey. *European Journal of Operational Research*, 290(3), 807-828.
10. Gambella, C.; Ghaddar, B.; and Naoum-Sawaya, J. (2021). Optimization problems for machine learning: A survey. *European Journal of Operational Research*, 290(3), 807-828.
11. Zhang, Y.; O'Neill, Z.; Dong, B.; and Augenbroe, G. (2015). Comparisons of inverse modelling approaches for predicting building energy performance. *Building and Environment*, 86(1), 177-190.
12. Bergh, D.V.D.; Clyde, M.A.; Gupta, A.R.K.N. (2021). A tutorial on Bayesian multi-model linear regression with BAS and JASP. *Journal of Behaviour Research Methods*, 53(6), 2351–2371.

13. Gebriel, Y.F. (2021). Bayesian regression model with application to a study of food insecurity in household level: A cross sectional study. *Journal of BMC Public Health*, 21(1), 1-10.
14. Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14(1), 045-076.
15. Hullman, J; and Gelman, A. (2021). Designing for interactive exploratory data analysis requires theories of graphical inference. *Journal of Harvard Data Science Review*, 3(3), 1-36.
16. Piter, W.; Loeneto, B.A.; and Jaya, P.H. (2018). Correlation between students' preferences on their teachers' code-switching and reading comprehension performance. *Journal of English Literacy Education*, 5(2), 182-193.
17. Chen, Y.; Tao, J.; Wang, J.; Chen, X.; Xie, J.; Xiong, J.; and Yang, K. (2019). The novel sensor network structure for classification processing based on the machine learning method of the ACGAN. *Sensors*, 19(14), 3145-3150.
18. Théate, T.; Wehenkel, A.; Bolland, A.; Louppe, G.; and Ernst, D. (2023). Distributional reinforcement learning with unconstrained monotonic neural networks. *Neurocomputing*, 534(1), 199-219.
19. Goonjur, M.K.; Sumitra, I.D.; and Supatmi, S. (2020). Enhanced the weighted centroid localization algorithm based on received strength signal in indoor wireless sensor network. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 1(1), 13-22.