IOP

by Hendra Bunyamin

PAPER • **OPEN ACCESS**

Classical and Deep Learning Time Series Prediction Techniques in the Case of Indonesian Economic Growth

To cite this article: Hendra Bunyamin and Meyliana 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1077** 012014

IOP Conf. Series: Materials Science and Engineering 1077 (2021) 012014 doi:10.1088/1757-899X/1077/1/012014

Classical and Deep Learning Time Series Prediction Techniques in the Case of Indonesian Economic Growth

Hendra Bunyamin[1] **and Meyliana**[2]

[1]Informatics Engineering, Maranatha Christian University, Bandung, 40164
[2]Accounting Department, Maranatha Christian University, Bandung, 40164

E-mail: `hendra.bunyamin@it.maranatha.edu`, `meyliana_oey@yahoo.com`

Abstract. Gross Domestic Growth (GDP) as a proxy of an economic growth in one country is defined as the market value of services and goods produced by the country for a year. As the GDP of one country gets higher and higher through the years, this indicator points out that in general the economic growth of the country is growing; specifically, the values of services and goods accumulated in the market are increasing. To determine what indicators of a country that affects one's economic growth remains an open question. Therefore, this research attempts to study those indicators and particularly utilize them to predict the economic growth. To answer those questions, this research employs diverse time series techniques ranging from classic time series analysis to machine learning and deep learning. Subsequently, our dataset comprises World Development Indicators (WDI) of Indonesia from 1962 to 2016. By measuring Root-Mean-Square Error (RMSE), we show Seasonal Autoregressive Integrated Average (SARIMA) and Convolutional LSTM give the the best performance from classical and deep learning techniques respectively. Our analysis shows SARIMA's performance is boosted by its ability to capture trend and seasonality of the dataset; equally important, Convolutional LSTM equipped with convolutions as part of reading input into LSTM units significantly boost the performance over either Convolutional or LSTM networks. Furthermore, our analysis points out that the indicator which most strongly contribute to predict GDP is CO_2 emission. This result agrees to the fact that some countries with high CO_2 emission also has high GDP as well; also this finding should warn Indonesian government of the increasing CO_2 pollution.

1. Introduction

Research on sources of long-run economic growth and development has always been an interesting topic among economists. At the same time, there is still a huge gap between developed and developing countries based on observed patterns in economic growth and time across countries. This phenomenon drives economists to conduct more qualitative and quantitative research to find explanation of the phenomenon with the latest information and mathematical methods [1].

Why does economic growth provide such a valuable information for a country? The Organization for Economic Co-operation and Development (OECD) [2] states that economic growth represents a significant instrument to decrease poverty and improve quality of society lives in developing countries. Furthermore, economic growth studies of a single country and comparing multiple countries have shown that an increasing economic growth leads to an accelerated Millenium Development Goals (MDGs) achievement. The eight United Nations MDGs are related to one another and designed to meet the needs of the world's poorest. Concretely, eight MDGs declared by United Nations in 2015 are described in Figure 1.

Figure 1. Eight MDGs designed to achieve a better and sustainable future for all countries in 2015

Specifically, Rodrik [3] states an increasing economic growth shall create a circle of chances and prosperity; in other words, an accelerated increasing economic growth will open enough intensive jobs opportunities for parents to provide education funding for their children. This conducive circumstance encourages the emergence of entrepreneurs with new openings which, in the end, drives a higher economic growth than before.

Moreover, in line with statement from Rodrik, learning from the history of one country's economic growth has profound insights, should everyone study country development growth. Economic growth can enable ones to learn a decent living both for themselves and their family, including for those who are in the lowest level. Concretely, economic growth can eliminate poverty; moreover, previous studies have also come to the same conclusion. Therefore, economic growth information should be important to be known by a country.

Unfortunately most studies related to economic growth utilize simple linear regression of exogenous factors without involving endogenous factors. Whereas Kibritcioglu and Dibooglu [1] state economic growth is too complex to be analyzed only by simple linear regression, firstly this research is expected to be able to answer the information needs regarding the calculation of predictable economic growth for the future by using time series methods [4, 5, 6, 7, 8] and machine learning [9, 10, 11, 12] to analyze and predict time series. Secondly, this research also analyzes world development indicators (WDIs) that influence economic growth of a country, specifically of Indonesia. To the best of our knowledge, this article presents the first exploration on predicting Indonesian economic growth using classical and deep learning time series techniques. Moreover, this article provides several baselines for predicting Indonesian economic growth problem.

Our source code is published at `https://github.com/hbunyamin/hibah_internal-2019-world_development_indicators-codes`.

2. Literature Review

Related work of this research is divided into three parts as follows: functionality decomposition in temporal data mining, time series forecasting techniques, and deep learning techniques.

2.1. Functionality decomposition in temporal data mining

Temporal data is a type of data which varies through time and indexed by a time component. Pattern extraction processes from temporal data have been studied extensively in a work by Mahalakshmi et al. [13]. In short, functionality decomposition can be distinguished into temporal classification, temporal prediction, temporal association, temporal clustering, temporal regression, temporal summarization, and temporal outlier detection. Our research is categorized as temporal prediction; specifically, we look for a GDP model prediction with country indicators as attributes.

2.2. Time series forecasting techniques

In general, regression techniques can be applied into predicting temporal data. These techniques are capable of modeling various relationships between dependent variables and independent variables where, in our case, the dependent variable is GDP and the independent variables are world development indicators (WDI) which affect the GDP. Categorically, [8] describe that there are three types of regression techniques. Firstly, the regression technique which models the relationship between a dependent variable and independent variables without any time component. This model is commonly named an explanatory model [14] and depicted as

$$y = f(x_1, x_2, \dots, x_n, \text{error}) \tag{1}$$

with y is an independent variable, x_1, x_2, ..., x_n are dependent variables. Specifically, error in Equation (1) denotes errors caused by random variation and other relevant variables which are not accommodated in the equation. In machine learning (ML), a dataset is divided into train and test set and a model which is similar to Equation (1) is trained on the train set and tested on the test set; therefore, the equation can also be a prediction model. Our research utilizes this regression type as machine learning models such as Ridge Regression [15], Lasso Regression [16, 17], Multilayer Perceptron [18], Random Forest [19, 20], and XGBoost [21]. Moreover, deep learning models which are more sophisticated than the ML ones, such as Convolution Neural Network (CNN) [22, 23, 24] are also employed for GDP prediction in this research.

The second type is a regression technique which takes into account the time component as described as follows:

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, y_{t-3}, \dots, \text{error}), \tag{2}$$

with y_{t+1}, y_t, and y_{t-1} are the value of independent variable at time $t+1$, t, and $t-1$ respectively. The error in Equation (2) is the same as the one in Equation (1). Particularly, Equation (2) is commonly named a univariate time series model. Classical models employed in this research are simple mean, median, and persistence methods, and exponential smoothing for time series (ETS) [25, 26]. Additionally, deep learning techniques for this type which are utilized are Long Short-Term Memory (LSTM) [27] and CNN-LSTM [27, 22, 25, 28].

Finally, the third type regression technique is the one which combines features from the first and second regression types. Accordingly, the third regression type for our problem can be written as

$$y_{t+1} = f(y_t, x_1, x_2, \dots, x_n, \text{error}). \tag{3}$$

where y_{t+1} and y_t are independent variables at time $t+1$ and t respectively. We also employ this regression type which represented by Seasonal Autoregressive Integrated Moving Average (SARIMA) [8, 26, 29] and a combination of CNN and LSTM (ConvLSTM) [22, 25, 27]. Specifically, SARIMA has the capability to model univariate time series with other features such as trend and seasonal components and ConvLSTM is a deep learning technique where the LSTM units of this model are modified into CNNs for reading input data.

IOP Conf. Series: Materials Science and Engineering 1077 (2021) 012014 doi:10.1088/1757-899X/1077/1/012014

3. Research Methodology

In general our approach consists of four steps [25] as follows: preprocessing the WDI dataset and designing experimental setup.

3.1. Preprocessing the WDI dataset

The Indonesian WDI dataset is divided into two parts that are the world development indicators (X) and the GDP (Y). The first part contains of $1,329$ world development indicators (https://data.worldbank.org/indicator) where each indicator is a time series and the second part comprises 9 GDPs of Indonesia. Both parts have the same time span from year 1960 to 2016.

The preprocessing on the dataset has been done as follows: indicator time series whose information are incomplete through the time span are removed and time series that are white noise are also removed. Concretely, white noise time series are sequences of random number; therefore, they are unpredictable. We determine whether or not a sequence is white noise by checking the autocorrelation of the time series. Autocorrelation (a_k) computes the linear relationship between lagged values (x_t and x_{t-k}) in a time series with k is a parameter defining the lag between two consecutive values [8]. The value a_k is defined as

$$a_k = \frac{\sum_{t=k+1}^{T}(x_t - \overline{x})(x_{t-k} - \overline{x})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}, \tag{4}$$

where T is the number of time components in the time series. Furthermore, the autocorrelations of a white noise time series for different k are approximately zero; however, not all values of autocorrelations are zeros because of random variation. Specifically, more than 95% of the autocorrelation values are within the range of $\pm\frac{2}{\sqrt{T}}$. Figure 2 shows an example of a white noise time series.

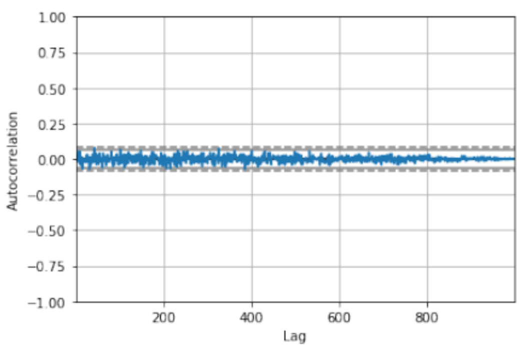

Figure 2. Autocorrelation of white noise time series (y). Both blue dotted lines represent $\pm\frac{2}{\sqrt{T}}$ and the autocorrelation values are within those values

On the other hand, Figure 3 shows the GDP time series of Indonesia is not white noise because less than 5% of autocorrelation values are within the range of $\pm\frac{2}{\sqrt{T}}$.

After white noise and incomplete world development indicators are removed, the number of indicators drops to 163 in a time span of 52 years. Additionally, in the second part of the dataset, there are nine different measurement metrics of GDP which are basically the same. We opt for GDP per capita with constant 2010 US$ inasmuch as it fairly reflects one country's economic

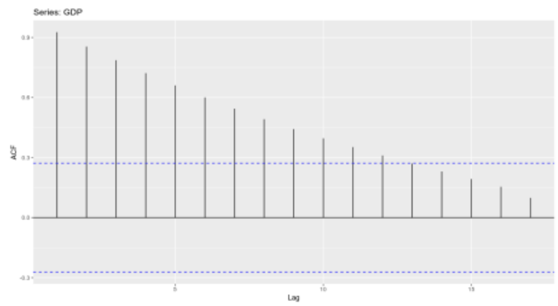

Figure 3. Autocorrelation of Indonesian GDP from year 1962 to 2013

output by dividing GDP by number of its people; moreover, the constant value of the currency makes GDP comparable during the time span of our experiment.

3.2. Designing experimental setup
Firstly all our prediction models are set up and summarized in Table 1. Specifically, we combine all classical, machine, and deep learning techniques.

Table 1. Classical, machine learning, and deep learning techniques from various different resources

Classical Techniques	Deep Learning Techiques
Simple mean, median [26, 29]	Ridge regression [15], Lasso regression [16, 17]
Simple persistence [8, 26, 29]	Random Forest [19], XGBoost [21]
Exponential smoothing (ETS) [8, 26, 29, 30]	Multilayer Perceptron [18]
SARIMA [8, 26, 29]	CNN [22], LSTM [27]
	CNN-LSTM, ConvLSTM [22, 25, 27]

After arranging the prediction models, we design an experimental setup to answer our first research question as follows:

- Dividing the dataset into train and test set. The train set consists of instances from year 1962 to 2003 (42 years) and the test set contains instances from year 2004 to 2013 (10 years).
- Train classical, machine learning, and deep learning models on the train set.
- Evaluate classical models with walk-forward validation. Walk-forward validation is a common practice in time series forecasting as it increases the number of train instances for models to learn at each time step [26].
- Evaluate machine and deep learning models on the test set.
- Both classical, machine learning, and deep learning models are measured by root mean squared-error (RMSE) evaluation metric.

Next, the second experimental setup aims to answer the question on the relationship between world development indicators and GDP. In this setup, we employ full dataset without any division and model both linear and non-linear relationships by multiple regression with Lasso regularization technique [31] and Random Forests [19].

4. Results

This section consists of three subsections which are results of GDP prediction with classical methods, deep learning methods, and finding relationships between world development indicators and GDP.

4.1. GDP prediction with classical methods

This article utilizes several classical time series forecasting methods as follows: Simple method, Exponential Smoothing (ETS), and Seasonal Autoregressive Integrated Moving Average (SARIMA). Table 2 describes hyperparameters and the RMSE results of these classical methods.

Table 2. Classical methods accompanied by their optimized hyperparameters

Methods	Hyperparameters
Simple method	config=`persist`, offset=1 & length=1
Simple method	config=`mean`, offset=1 & length=2
Simple method	`median`, offset=1 & length=2
ETS method	trend=`mul`, damped=True, seasonal=`None`, seasonal_periods=`None` use_boxcox=True, remove_bias=False
SARIMA method: Trend	order=`(1,1,0)`, seasonal_order=`(2,0,1,0)`, trend=`'t'`
SARIMA method: Trend & Seasonality	order=`(0,1,2)`, seasonal_order=`(0,0,1,12)`, trend=`'t'`

Table 3. GDP prediction results by classical methods measured by RMSE (the lower RMSE equals to the better performance of the model)

Methods	RMSE
Simple method	125.7454
Simple method	184.9804
Simple method	184.9804
ETS method	28.4467
SARIMA method: Trend	20.5524
SARIMA method: Trend & Seasonality	18.0193

In order to find hyperparameters that give the lowest RMSE of the models, we do grid search process on each time series model. Concretely, for simple method, the grid search process

IOP Conf. Series: Materials Science and Engineering 1077 (2021) 012014 doi:10.1088/1757-899X/1077/1/012014

takes values of `mean`, `median`, and `persist` with `length` from 1 to 42. ETS method grid search process searches for optimized hyperparameters among trend=`['add', 'mul', None]`, damped=`[True, False]`, seasonal=`['add', 'mul', None]`, use_boxcox=`True, False`, remove_bias=`True, False`. Ultimately, SARIMA method has three hyperparameters which govern nature of the model, trend, and seasonality [25]. Specifically, the grid search process scans for the most optimized hyperparameters among

(i) `order`: a tuple of parameters which contains parameters (`p`, `d`, and `q`) to model the trend,

(ii) `seasonal_order`: a tuple which holds parameters (`P`, `D`, dan `Q`) to model seasonality,

(iii) `trend`: a parameter controlling the trend as one of these options: `'n'` (no trend), `'c'` (constant trend), `'t'` (linear trend), and `'ct'` (constant linear trend).

Grid search process done on SARIMA method covers hyperparameters as follows: `order`={`p=[0,1,2], d=[0,1], q=[0,1,2], t=['n','c','t','ct']`}, `P=[0,1,2], D=[0,1], Q=[0,1,2]`. After grid processes are done thoroughly, Table 2 finally shows that the lowest RMSE is achieved by SARIMA which considers Trend & Seasonality. Moreover, this combination of Trend & Seasonality of SARIMA models GDP Indonesia better than other classical methods as the more hyperparameters a model has, the better its performance is. This finding and M4 [32], a competition whose purpose is to advance predictions accuracy in theory and applications of time series analysis, are in accord.

4.2. GDP prediction with deep learning methods

Deep learning models specifically designed to predict Indonesian GDP in this research are described in Table 4.

Table 4. List of deep learning techniques to predict GDP. Train set and test set are the GDP and indicators from year 1962 to year 2003 and from year 2004 to year 2013 respectively.

Deep Learning Models	Hyperparameters
MP	# neurons in hidden layer=500, # epoch=100

CNN	# filter=64, # kernel size=2,
	activation=relu, optimizer=adam

LSTM	# stack in LSTM=50, activation=relu
	optimizer=adam

CNN-LSTM	# filter=64, # stack in LSTM=50 # kernel size=2, activation=relu, optimizer=adam
ConvLSTM	# filter=64, # stack in LSTM=50 # kernel size=2, activation=relu, optimizer=adam

Table 5 shows that the Convolutional LSTM model gains the minimum value of RMSE. Convolutional LSTM method reads subsequence blocks of time series observations through convolutional processes similar to CNN method. Apparently, this subsequence blocks reading give an advantage for the model to predict because the CNN operation on each step contributes additional information enriching the model to predict better. Based on the RMSE results, it appears that the Convolutional LSTM still have difficulty in predicting Indonesian economic growth due to its large RMSE values. Nevertheless, those models in Table 4 may be initialized as baselines for continuing research to pursue better performance. Visually, Figure 4 shows the comparison of our proposed deep learning techniques.

IOP Conf. Series: Materials Science and Engineering 1077 (2021) 012014 doi:10.1088/1757-899X/1077/1/012014

Table 5. GDP prediction results from deep learning techniques measured by RMSE

Deep Learning Models	**RMSE ± Std**
MP	190.298 ± 0.000
CNN	195.300 ± 56.693
LSTM	593.214 ± 211.292
CNN-LSTM	300.009 ± 224.893
ConvLSTM	179.130 ± 9.001

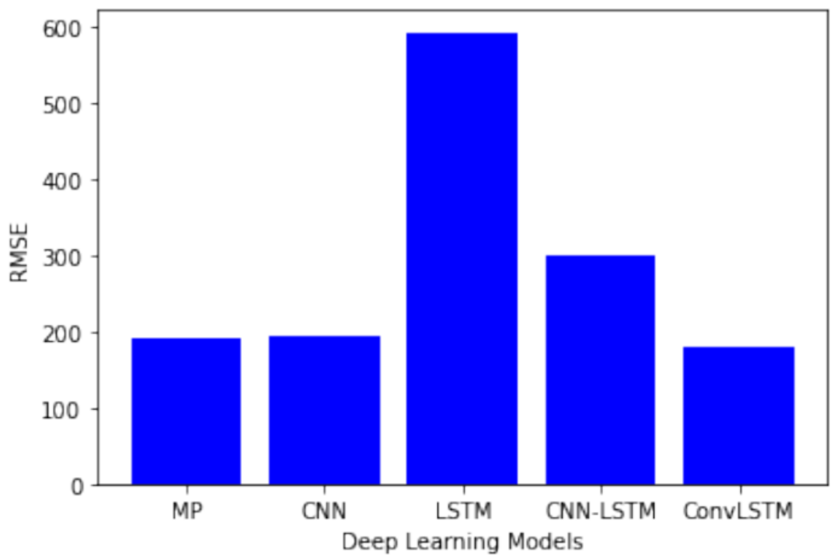

Figure 4. RMSEs from different deep learning techniques (the smaller the RMSE, the better the technique's performance)

4.3. Importance of world development indicators on predicting GDP

Our next analysis considers importance of the indicators on predicting GDP. In order to determine the importance of each indicator for predicting Indonesian GDP, we model the dataset by employing two models as follows:

- Random forest classifier [19] with number of tree estimator equals to 5. The importance of an indicator is calculated by the difference between impurity measures before reaching the indicator and at the indicator. Each calculation is weighted by its probability [33]. The more important an indicator is, the higher the difference value is.
- Lasso, a linear model designed to estimate sparse coefficients [16, 17] with $\alpha = 1$ and $\alpha = 0.01$. In this model, the importance of an indicator is measured by its coefficient [34]. The more important an indicator is the higher, the coefficient value is.

Ten most important indicators for predicting Indonesian GDP are shown in Table 6.

Interestingly, our research discovers CO_2 emissions are the most important indicator for predicting GDP. This finding suggests CO_2 emissions generated from heavy industries in Indonesia should be contributing the most to the GDP.

Table 6. Ten indicators which contribute the most for predicting Indonesian GDP

Ranking	World development indicators
1	CO_2 emissions
2	GNI (constant LCU)
3	GNI (current LCU)
4	GNI per capita growth (annual %)
5	GNI (constant 2010 US$)
6	General government final consumption expenditure (current LCU)
7	Net bilateral aid flows from DAC donors, United Kingdom (current US$)
8	Food production index (2004-2006 = 100)
9	Permanent cropland (% of land area)
10	GNI per capita (constant 2010 US$)

5. Conclusion

This research has modeled the growth of Indonesian economics represented by GDP from year 1962 to 2016. Moreover, we provide classical and deep learning time series models to predict the GDP. For the classical time series model, Seasonal Autoregressive Integrated Moving Average (SARIMA) gives the minimum RMSE; this finding is in accord with M4 Time Series competition [32]. Besides, Convolutional LSTM achieves the minimum RMSE for deep learning time series models as it takes an advantage of subsequence blocks reading to predict more accurate than other models. Lastly, CO_2 emissions are the most important world development indicator for predicting GDP. This finding can signify Indonesia relies on heavy industries to increase its GDP in coming years.

As to future work, we plan to study other potential signals which can be inferred from external sources for predicting economic growth.

References

[1] Kibritcioglu A and Dibooglu S 2001 Long-run economic growth: an interdisciplinary approach *Knowledge, Technology & Policy* **13** pp 59–70 ISSN 1874-6314 DOI: 10.1007/BF02693990
[2] OECD 2019 *Together, We Create Better Policies for Better Lives* Retrieved on 13 February 2020 from `https://www.oecd.org/about`
[3] Rodrik D 2008 *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth* (New York: Princeton University Press)
[4] Gilbert P D 1993 *State Space and ARMA Models: An Overview of the Equivalence* Working Paper 93-4 (Ottawa, Canada: Bank of Canada) Retrieved from http://www.bankofcanada.ca/1993/03/publications/research/working-paper-199/
[5] Gilbert P D 1995 Combining var estimation and state space model reduction for simple good predictions *J. of Forecasting: Special Issue on VAR Modelling* **14** pp 229–50

[6] Gilbert P D 2000 A note on the computation of time series model roots *Applied Economics Letters* **7** pp 423–4 DOI: 10.1080/135048500351096
[7] Gilbert P D 2006 *Brief User's Guide: Dynamic Systems Estimation* Retrieved from http://cran.r-project.org/web/packages/dse/vignettes/Guide.pdf
[8] Hyndman R J and Athanasopoulos G 2018 *Forecasting: Principles and Practice* (Melbourne: OTexts)
[9] Ahmed N, Atiya A, Gayar N and El-Shishiny H 2010 An Empirical Comparison of Machine Learning Models for Time Series Forecasting *Econometric Reviews* **29** pp 594–621
[10] Fakhrazari A and Vakilzadian H 2017 A survey on time series data mining *2017 IEEE Int. Conf. on Electro Information Technol.(EIT)* pp 476–481 ISSN 2154-0373
[11] Aungiers J 2018 *Time Series Prediction using LSTM Deep Neural Networks* Retrieved on 7 November 2018 from https://www.altumintelligence.com/articles/a/Time-Series-Prediction-Using-LSTM-Deep-Neural-Networks
[12] Brownlee J 2018 *Deep Learning with Time Series Forecasting, Machine Learning Mastery* Retrieved on 18 December 2018 from https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python
[13] Mahalakshmi G, Sridevi S and Rajaram S 2016 A survey on forecasting of time series data *2016 Int. Conf. on Computing Technologies and Intelligent Data Eng. (ICCTIDE'16)* (IEEE) pp 1–8
[14] Shmueli G *et al.* 2010 To explain or to predict? *Statistical Sci.* **25** pp 289–310
[15] Rifkin R M and Lippert R A 2007 *Notes on Regularized Least Squares* Retrieved from https://www.researchgate.net/publication/37997883_Notes_on_Regularized_Least_Square
[16] Kim S j, Koh K, Lustig M, Boyd S and Gorinevsky D 2007 An interior-point method for large-scale l1-regularized logistic regression *J. of Machine Learning Research* **8** pp 1519–55
[17] Friedman J, Hastie T and Tibshirani R 2010 Regularization paths for generalized linear models via coordinate descent *J. of Statistical Software* **33** pp 1
[18] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors *Nature* **323** pp 533
[19] Breiman L 2001 Random forests *Machine Learning* **45** pp 5–32
[20] Jiang N, Fu F, Zuo H, Xiuping Z and Zheng Q 2020 A municipal pm2.5 forecasting method based on random forest and wrf model *Eng. Letters* **28** pp 312–21
[21] Chen T and Guestrin C 2016 Xgboost: A scalable tree boosting system *CoRR* **abs/1603.02754** pp 1 (*Preprint* 1603.02754)
[22] Lecun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. of the IEEE* **86** pp 2278–324
[23] Bunyamin H, Heriyanto, Novianti S and Sulistiani L 2019 Topic clustering and classification on final project reports: a comparison of traditional and modern approaches *IAENG Int. J. of Computer Sci.* **46** pp 506–11
[24] Asriny D M, Rani S and Hidayatullah A F 2020 Orange fruit images classification using convolutional neural networks *IOP Conf. Series: Materials Sci. and Eng.* **803** pp 012020
[25] Brownlee J 2019 *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs, and LSTMs in Python* Retrieved on 27 May 2019 for https://machinelearningmastery.com/deep-learning-for-time-series-forecasting
[26] Brownlee J 2019 *Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future* Retrieved on 27 May 2019 from https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python/
[27] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Computation* **9** pp 1735–80
[28] Aspiras L L F, Hanbal I F, Ingosan J S and Panlilio K B P 2020 Establishing the syntactic rules of the kankanaey dialect using RNN *IOP Conf. Series: Materials Sci.ience and Eng.* **803** pp 012023
[29] Brownlee J 2018 *11 Classical Time Series Forecasting* Retrieved on 8 January 2019 from https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet
[30] NIST S *et al.* 2012 Engineering statistics handbook *E-Handbook of Statistical Methods* (NIST)
[31] James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning* vol 112 (New York: Springer)
[32] Makridakis S, Spiliotis E and Assimakopoulos V 2018 The m4 competition: results, findings, conclusion and way forward *Int. J. of Forecasting* **34** pp 802–8 ISSN 0169-2070
[33] Raschka S and Mirjalili V 2017 *Python Machine Learning Second Edition* (Mumbai: Packt Publishing Ltd)
[34] Hastie T, Tibshirani R and Friedman J 2008 *The Elements of Statistical Learning Second Edition* Springer Series in Statistics (New York: Springer New York Inc.)

IOP

ORIGINALITY REPORT

20%	17%	14%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	irep.iium.edu.my Internet Source	6%
2	hdl.handle.net Internet Source	1%
3	Andrzej Wodecki. "The Reserve Price Optimization for Publishers on Real-Time Bidding on-Line Marketplaces with Time-Series Forecasting", Foundations of Management, 2020 Publication	1%
4	Mridu Sahu, D.C. Jhariya, Richa Singh, Ishika Srivastava, Sunny Kumar Mishra. "Gis Based Spatial Analysis and Prediction of Covid-19 Cases", Journal of Physics: Conference Series, 2022 Publication	1%
5	www.coursehero.com Internet Source	1%
6	Jie Yang, Lin Zou, Yiying Wei, Pengju Yuan, Chen Zhou. "Health Status Prediction of	1%

	Lithium Battery Based on LSTM Model with Optimization Algorithms", Journal of Physics: Conference Series, 2023 Publication	
7	data.worldbank.org Internet Source	1 %
8	jatit.org Internet Source	1 %
9	repository.unisba.ac.id:8080 Internet Source	1 %
10	Submitted to Indiana University Student Paper	1 %
11	www.alkaline-ml.com Internet Source	<1 %
12	"Artificial Intelligence and Machine Learning in Healthcare", Springer Science and Business Media LLC, 2021 Publication	<1 %
13	dokumen.pub Internet Source	<1 %
14	www.proceedings.com Internet Source	<1 %
15	id.123dok.com Internet Source	<1 %

16 Weizhen Gao, Menggang Li, Chengzhang Zou. "Analysis of the Impact of ESG on Corporate Financial Performance under the Epidemic Based on Static and Dynamic Panel Data", Wireless Communications and Mobile Computing, 2022
Publication
<1%

17 api.worldbank.org
Internet Source
<1%

18 Ketaki Anandkumar Pattani, Sunil Gautam. "chapter 1 Introduction to Meteorology and Weather Forecasting", IGI Global, 2022
Publication
<1%

19 arrow.tudublin.ie
Internet Source
<1%

20 sciendo.com
Internet Source
<1%

21 valuesimplex.com
Internet Source
<1%

22 Mengqi Wu, Paula Branco, Janny Xue Chen Ke, David B. MacDonald. "Artifact Detection in Invasive Blood Pressure Data using Forecasting Methods and Machine Learning", 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020
Publication
<1%

23	digital.library.unt.edu Internet Source	<1%
24	jfds.pm-research.com Internet Source	<1%
25	theses.hal.science Internet Source	<1%
26	www.pure.ed.ac.uk Internet Source	<1%
27	www.yumpu.com Internet Source	<1%
28	Mohan Vamsi Nallapareddy, Rohit Dwivedula. "ABLE: Attention based learning for enzyme classification", Computational Biology and Chemistry, 2021 Publication	<1%
29	Sayed S R Moustafa, Sara S Khodairy. "Comparison of different predictive models and their effectiveness in sunspot number prediction", Physica Scripta, 2023 Publication	<1%
30	analyticsindiamag.com Internet Source	<1%
31	digitalcommons.wayne.edu Internet Source	<1%
32	iopscience.iop.org Internet Source	

		<1%
33	link.springer.com Internet Source	<1%
34	vdoc.pub Internet Source	<1%
35	www.arxiv-vanity.com Internet Source	<1%
36	www.foodandnutritionjournal.org Internet Source	<1%
37	www.frontiersin.org Internet Source	<1%
38	Jia Lu, Gang Li, Chuntian Cheng, Benxi Liu. "A long-term intelligent operation and management model of cascade hydropower stations based on chance constrained programming under multi-market coupling", Environmental Research Letters, 2021 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		