International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I6P139
Volume 73 Issue 6, 476-484, June 2025

Original Article

Forecasting Teaching Materials Using the Autoregressive Integrated Moving Average (ARIMA) Method Case Study: Universitas Terbuka

Bagus Arif Wicaksono[1], Tuga Mauritsius[2]

[1,2]Information System Management Department, Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia.

[1]Corresponding Author : bagus.wicaksono003@binus.ac.id

Received: 23 December 2024 Revised: 03 April 2025 Accepted: 10 June 2025 Published: 28 June 2025

Abstract - *Universitas Terbuka (UT) is an open and distance learning university in Indonesia, where teaching materials serve as the primary learning resources. Currently, UT relies on manual calculations using Microsoft Excel to estimate the required number of printed materials. This approach often results in inaccurate forecasts, leading to either a surplus or shortage of teaching materials. Such discrepancies negatively impact operational efficiency and cost-effectiveness. Repeated duplication increases unit costs, while surplus materials risk expiration and waste due to content updates or limited shelf life. This study applies a machine learning approach, specifically the ARIMA method, within the CRISP-DM framework to accurately forecast teaching material needs. Results show that the ARIMA (2,1,0) model, with parameters p=2, d=1, q=0, and AIC=34.512 and BIC=30.591, provides the best performance. Teaching material EKMA4434, used in 11 study programs, recorded the lowest RMSE of 5,909.881 units, indicating an average prediction error of that magnitude against actual usage.*

Keywords - *Forecasting, Machine Learning, Time series, Teaching material, Arima.*

1. Introduction

Universitas Terbuka (UT) is a public university in Indonesia that implements a distance learning system, serving over 520,000 students. At UT, teaching materials are the primary learning resources and must be provided to all students. With such a large student population, the demand for teaching materials continues to grow. Currently, UT manages approximately 1,100 titles of teaching materials distributed across 34 academic programs. The Pusat Pengelolaan Bahan Ajar (Puslaba), or the Center for Teaching Material Management, procures and distributes these materials. Each semester, Puslaba handles millions of copies. For instance, according to the 2023 State-Owned Goods Balance Sheet Report, Puslaba managed 3,858,195 copies of teaching materials. To support this large volume, Puslaba utilizes an inventory management system called Sistem Informasi Tiras dan Transaksi Bahan Ajar (SITTA), which is specifically designed to teach material stock and transaction management.

Currently, Puslaba estimates the number of teaching materials to be duplicated manually using Microsoft Excel, relying solely on the number of students from the previous semester (N-1) as a basis for projection. This manual method often leads to inaccuracies in demand forecasting, resulting in either a surplus or shortage of teaching materials. Consequently, the delivery of materials to students becomes operationally inefficient, and the associated costs are suboptimal. In cases of shortages, repeated duplications lead to increased procurement costs. Conversely, excess inventory carries the risk of disposal, as teaching materials at Universitas Terbuka have limited usage periods and may become obsolete or inactive. Data from 2023 indicates repeated additional procurements were required because initial forecasts underestimated actual demand, necessitating reprints to meet the needs of the current semester.

Table. 1 Data on teaching material needs periode 2023

No	Code BA	Procurement at the Beginning of the Semester	Additional Needs
1	EKMA4159	3.500	465
2	ESPA4111	12.300	5.207
3	EKMA4434	2.200	1.596
4	MKDU4109	25.400	5.666
5	ISIP4112	12.800	3.422

Furthermore, between 2021 and 2024, several teaching materials were discarded due to overestimation in previous forecasts and subsequent changes in editions, rendering the existing stock obsolete or inactive and requiring disposal.

Table. 2 Data on destruction of teaching materials

Years	Number Of Titles	Quantity Books (Eks)	Amount Weight (Kg)	Sum Of Cost (Rp)
2021	697	630.872	68.540	7.000.353
2023	355	143.419	41.288	2.681.928
2024	1.280	181.934	56.047	3.625.027

1.1. Research Gap

Previous studies have indicated that accurate forecasting of material demand must be aligned with implementing appropriate machine learning technologies and accompanied by performance evaluation of the forecasting results. The forecasting method adopted in this study is the Time Series Moving Average, specifically utilizing the Autoregressive Integrated Moving Average (ARIMA) model. The selection of the optimal ARIMA model is essential to achieving high accuracy, which is assessed based on actual usage data of teaching materials in each academic semester. This approach aims to ensure that the forecasted demand can effectively support the provision of teaching materials for Universitas Terbuka students. In response to these issues, this study aims to examine the application of machine learning techniques to forecast the duplication needs of teaching materials at Universitas Terbuka and to evaluate the performance of the forecasting results produced by the selected machine learning model.

2. Literature Review

According to Prof. Atwi Supratman and Prof. Aminudin Zuhairi (2004), distance education relies heavily on instructional media-such as printed materials, audiovisual content, and electronic resources-rather than conventional face-to-face teaching methods. These media contain educational content or academic scripts specifically designed for distance learning. Interaction between educational administrators and students is also facilitated through these media. Operationally, distance education resembles an industrial process involving mass-scale systems such as the production and printing of teaching materials, distribution of course modules, management of course registration documents, and the use of communication and information technologies to support both academic and administrative services.

In relation to the problem addressed in this study, relevant work was conducted by Tang, Y.-M., Ho, G.T.S., Lau, Y.-Y., and Tsui, S.-Y. (2022), which examined the Integrating smart warehouse and production management with demand forecasting in small-scale cyclical industries. Earlier studies by Box et al. (2016), Holt (1957), and Winters (1960) investigated classical forecasting methods based on chronological time series data, aiming to predict future trends by extrapolating statistical information derived from historical data. Forecasting methods are generally divided into two main approaches: univariate and multivariate. The univariate approach relies solely on time series data, while the multivariate approach incorporates external variables. Two of the most common univariate forecasting methods are Exponential Smoothing (ES) and Autoregressive Integrated Moving Average (ARIMA). Sri Baginda Dalimunthe, Sukaria Sinulingga, and Rosnani Ginting (2023) conducted a study exploring the application of machine learning for demand forecasting across various industrial sectors, from small- to large-scale enterprises. Their discussion covers machine learning models, data processing techniques, and research variables to compare the accuracy of different approaches used in forecasting. Hassyddiqy and Hasdiana (2023) analyzed sales forecasting using the ARIMA method to predict future clothing production needs at Huebee Indonesia, aiming to improve production and sales planning.

Xing Yee Leong, Nethal Jajo, Shelton Peiris, and Mohamed Khadra (2023) applied a hybrid ARIMA–machine learning model to forecast elective surgery demand. Their research compared ARIMA with other machine learning and deep learning models, such as ANN, LSTM, and Random Forest, and found that the ARIMA–ANN model yielded the best performance. Adeline P. Dela Cruz and Ma. Leslie B. Basallo (2020) conducted a study to forecast student enrollment for academic years 2019–2020 through 2024–2025, using historical enrollment data as the basis for prediction. In the present study, forecasting techniques are essential to estimate the quantity of teaching materials needed effectively and efficiently.

Forecasting approaches can be broadly categorized into two types:

1. Qualitative approaches are based on subjective judgment and include consumer surveys, sales force composites, expert opinion, and the Delphi method.
2. Quantitative approaches rely on statistical methods like time series and causal models.

These forecasting approaches are classified in Figure 1 (Tersine, 1994).

Time series methods rely on historical data collected over a defined time period (weekly, monthly, or yearly) and focus on identifying patterns within the data. In contrast, causal methods are used to identify correlations between independent and dependent variables, such as through linear regression analysis. The ARIMA method is a widely used statistical approach for time series forecasting. It combines three main components: autoregression (AR), differencing for integration (I), and moving average (MA). This method is effective in modeling non-stationary time series data by transforming it into a stationary format using differencing (Box & Jenkins, 1976). Furthermore, ARIMA is a foundational model for various modern hybrid models integrating statistical techniques with machine learning (Zhang, 2003).

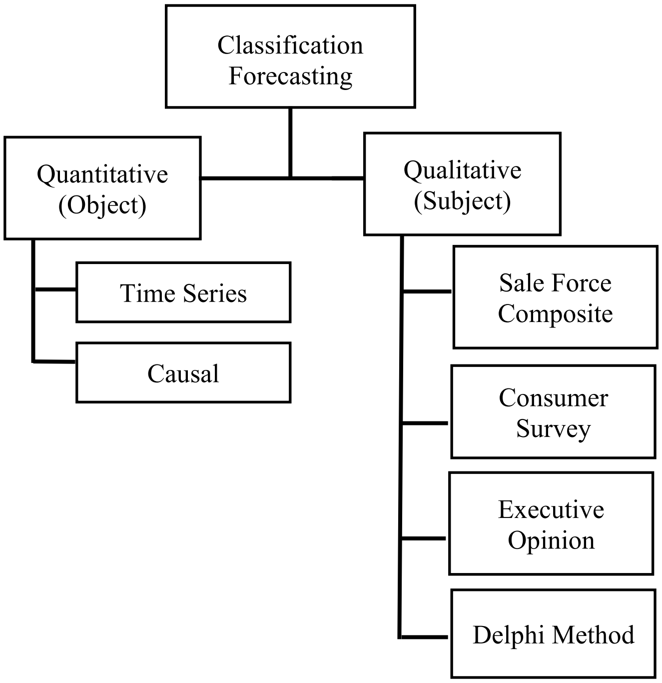

Fig. 1 Forecasting classification (Tersine, 1994) [16]

Dela Cruz and Basallo (2020) used ARIMA to analyze time series data by testing various combinations of parameters (p, d, q) and selecting the model with the lowest Akaike Information Criterion (AIC) value as the best fit. When applying the ARIMA model, a balance must be maintained between model fit and complexity. This is where AIC and the Bayesian Information Criterion (BIC) are critical; they help evaluate and compare models to select the most accurate and efficient option. The model with the lowest AIC value is considered the most suitable.

To evaluate forecasting performance, this study uses three key metrics:

- Mean Absolute Error (MAE): the average absolute differences between predicted and actual values (Zhang et al., 2011).
- Mean Squared Error (MSE): the average of the squared differences between predicted and actual values. A lower MSE indicates better model performance.
- Root Mean Squared Error (RMSE): the square root of MSE, which restores the original units of the data and emphasizes larger errors.

Makridakis et al. (2018) emphasized the importance of using MSE and RMSE as evaluation metrics when comparing forecasting models, as these metrics account for the magnitude of error. This study adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, a widely accepted framework for data mining and predictive analytics frequently used by researchers to address complex data problems.

3. Proposed Methodology

To forecast the duplication needs of teaching materials using machine learning, this study employs a Time Series method with the ARIMA model. The forecasting results are evaluated using three performance metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). This research adopts the CRISP-DM methodology, which is outlined in the following stages:

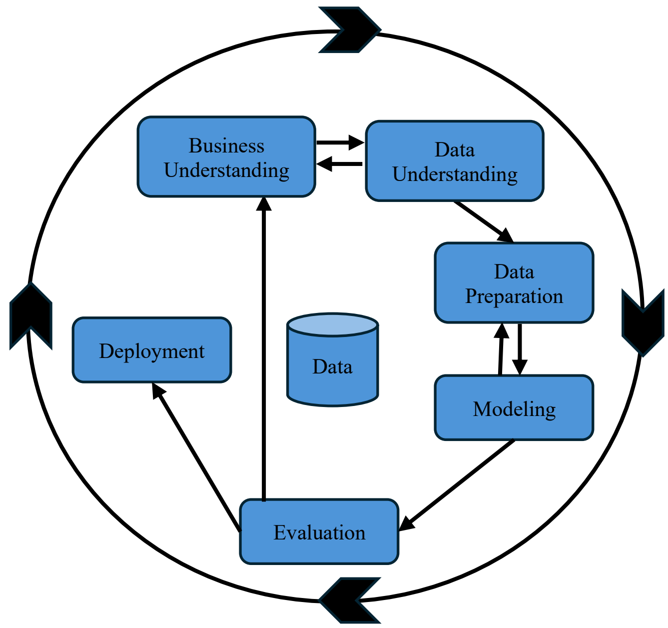

Fig. 2 CRISP-DM Stages [11]

3.1. Business Understanding

This stage provides an understanding of Universitas Terbuka's (UT) business processes in delivering higher education services to the wider Indonesian population. UT is a state university that implements an open and distance learning system, where students and lecturers are not required to be in the same place or time. This learning model demands that students become independent learners, with printed modules or teaching materials as the primary learning resources instead of face-to-face instruction. Therefore, the availability of these materials is essential and must be ensured by the university.

3.2. Data Understanding

Various datasets are needed to estimate the number of teaching materials required for duplication. These include student registration data, usage data of teaching materials for each academic program, the lifespan of materials, and other supporting information related to teaching materials used at Universitas Terbuka.

3.3. Data Preparation

In the data collection process for forecasting the demand for teaching materials, it is crucial to ensure the data is reliable, valid and can be processed into actionable information for forecasting purposes. The following are the stages involved in preparing the dataset for forecasting:

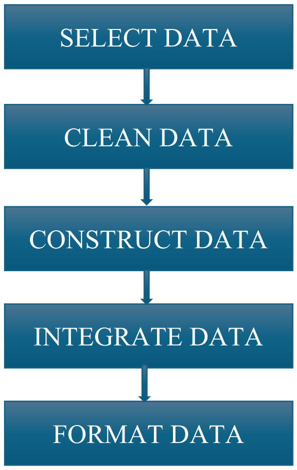

Fig. 3 Data preparation stages

3.3.1. Select Data

The dataset obtained from Universitas Terbuka consists of spreadsheet files containing cohort data on student enrollment per semester, per study program, and regional office (UT Daerah), covering semesters from 2018 (Odd) to 2023 (Even). The demand for teaching materials is determined based on materials listed in the curriculum for each study program and semester.

3.3.2. Clean Data

The collected data must be cleaned to eliminate inconsistencies or extreme outliers. For example, data on student enrollment showing unusually large increases or decreases from the semester average are excluded to improve forecast accuracy.

3.3.3. Construct Data

This process involves deriving additional attributes from the primary data. For example, from the student enrollment data per study program per semester, information on the corresponding teaching materials can be derived to calculate the total required quantities.

3.3.4. Integrate Data

Data from various sources must be integrated to provide a comprehensive dataset. In this study, student registration data from the 2018 odd semester to the 2024 even semester is merged to identify trends in enrollment increases or decreases, which form the basis for forecasting teaching material needs.

3.3.5. Format Data

Since the data comes in various formats (e.g., DOC, PDF, XLS), standardization is necessary before loading it into a database or data processing tool. The formatted and harmonized data can then be visualized to generate insights required by stakeholders.

3.4. Modelling

This study uses a machine learning-based modeling technique with the ARIMA time series forecasting method. The model's accuracy is tested using MAE, MSE, and RMSE to determine the most effective and efficient approach for forecasting the number of teaching materials to be duplicated. The models are built, tested, and evaluated using Python-based data processing tools.

3.5. Evaluation

The following metrics are used in the evaluation process:

- Mean Absolute Error (MAE)
  The average of the absolute differences between actual and predicted values.
- Mean Squared Error (MSE)
  The average of the squared differences between the actual and predicted values. This metric penalises larger errors, making it sensitive to outliers.
- Root Mean Squared Error (RMSE)
  The square root of MSE brings the error metric back to the original unit of measurement, making it easier to interpret while emphasizing larger errors.

The study's results will be evaluated to determine whether the forecasting model can contribute to cost efficiency at Universitas Terbuka. The overall estimation process for teaching material duplication will also be reviewed to assess whether it yields effective and efficient quantity planning.

4. Results and Discussion

4.1. Business Understanding

To achieve optimal outcomes in its operations, Universitas Terbuka (UT) must implement a well-structured process to forecast the duplication needs of teaching materials. This involves thoroughly analyzing and processing existing data to generate actionable insights that can serve as a reference for material procurement decisions. An initial assessment of the appropriate tools and techniques is crucial to support this research. Several preparatory steps must be undertaken, including identifying the necessary data to produce the required information, selecting suitable tools for data processing, and determining the forecasting method that will be applied. These steps ensure the generation of valid and effective data to support decision-making.

4.2. Data Understanding

The data required to forecast teaching material duplication needs for supporting student learning at UT are sourced from various databases, including:

1. Student Enrollment History from UT's Student Record System Database.
2. Teaching Material Sales Data from the UT Teaching Material Management Database, collected during each registration period.

3. Teaching Material Procurement Data from the UT Teaching Material Management Database for each registration period.
4. Teaching Material Distribution Data sourced from the same database based on each semester's distribution records.
5. Teaching Material Inventory Data from UT's inventory system.

Since the mandatory purchase of teaching materials was implemented starting from Semester 2019.2, non-package students can register for courses freely each semester.

4.3. Data Preparation

Formatted data from various sources are integrated and processed to form a comprehensive dataset that stakeholders can visualise and utilise.

Key datasets include:

1. Teaching Material Usage per Course per Study Program

This dataset was derived from curriculum catalogues and details the use of teaching materials per course per semester. It enables the calculation of precise demand based on student numbers per study program and semester.

Table. 3 Teaching material data per course per program

No	Code BA	Name Of Program Study	Sem
1	EKMA4434	Ekonomi Pembangunan	8
2	EKMA4434	Manajemen	3
3	EKMA4434	Agribisnis Pertanian	6
4	EKMA4434	Agribisni Peternakan	6
5	EKMA4434	Agribisnis Perikanan	8
6	EKMA4434	Akuntansi	4
7	EKMA4434	Teknologi Pendidikan	6
8	EKMA4434	Sistem Informasi	1
9	EKMA4434	SAINS DATA	1
10	EKMA4434	Ekonomi Syariah	6
11	EKMA4434	Akuntansi Keuangan Publik	4

2. Grouped Teaching Material Usage

This dataset categorizes materials based on how many programs use them, which is used for forecasting model trials per material group.

Table. 4 Data on the classification of teaching material usage

Number Of Program Study	Teaching Material
15	ISIP4112,ISIP4215,ISIP4216
13	MKDK4002,MKDK4005,MKDU4109
12	IDIK4007,IDIK4500,ISIP4111
11	EKMA4434,IDIK4008,PTAP5406
10	BIOL4110,ESPA4122,ISIP4500,MKDK4001
9	EKMA4116,EKMA4214,IDIK4012,IDIK4017
8	EKMA4115,IDIK4009,IDIK4010,IDIK4013
7	EKMA4111,ESPA4123,ISIP4130,ISIP4212
6	EKMA4158,ISIP4131,ISIP4211,MAPU5101
5	ADBI4201,ADPU4442,BIOL4417,EKMA5101
4	ADBI4211,ADPU4218,ADPU4330,ADPU4334
3	ADBI4330,ADBI4532,ADPU4331,ADPU4332
2	ADBI4235,ADBI4335,ADBI4336,ADBI4410
1	ADBI4130,ADBI4331,ADBI4333,ADBI4433

3. Student Enrollment per Registration Period per Study Program

This dataset formats student cohort information to estimate teaching material needs per registration cycle and semester.

Table. 5 Data sum of students per program per semester

Periode	Code Program Study	Name Program Study	Semester	Sum Of Student
20231	252	Sistem Informasi	8	522
20231	252	Sistem Informasi	6	894
20231	252	Sistem Informasi	2	2707
20231	252	Sistem Informasi	7	207
20231	252	Sistem Informasi	3	977
20231	252	Sistem Informasi	5	643
20231	252	Sistem Informasi	4	1583
20231	252	Sistem Informasi	1	3269

4. Teaching Material Distribution per Semester

The data on the number of teaching materials distributed per registration period results from formatting and compiling data from several sources, primarily derived from the usage data of teaching materials across all academic programs each semester. This dataset is essential for forecasting teaching material needs using the ARIMA time series model based on the volume of materials distributed. The dataset is presented as follows:

Table. 6 Data teaching material expenditure by period

Periode	Code BA	Sum Of BA
1-Jan-2018	EKMA4434	5.807
1-Jul-2018	EKMA4434	3.516
1-Jan-2019	EKMA4434	10.510
1-Jul-2019	EKMA4434	5.084
1-Jan-2020	EKMA4434	11.419
1-Jul-2020	EKMA4434	6.924
1-Jan-2021	EKMA4434	12.550
1-Jul-2021	EKMA4434	9.991

1-Jan-2022	EKMA4434	18.142
1-Jul-2022	EKMA4434	12.495
1-Jan-2023	EKMA4434	19.068
1-Jul-2023	EKMA4434	13.508
1-Jan-2024	EKMA4434	20.082

4.4. Modelling

In this study, the modeling technique utilizes a machine-learning approach. The method employed to forecast the duplication requirements of teaching materials is the Time Series method using the ARIMA model.

The model was developed, tested, and evaluated using data processing tools in Python via Google Colab. Google Colab (Collaboratory) is a cloud-based service provided by Google that allows users to write and execute Python code directly through a web browser.

For the forecasting model trial, the teaching material distribution dataset was divided into two parts:

1. Training Data 80%
2. Testing Data 20%

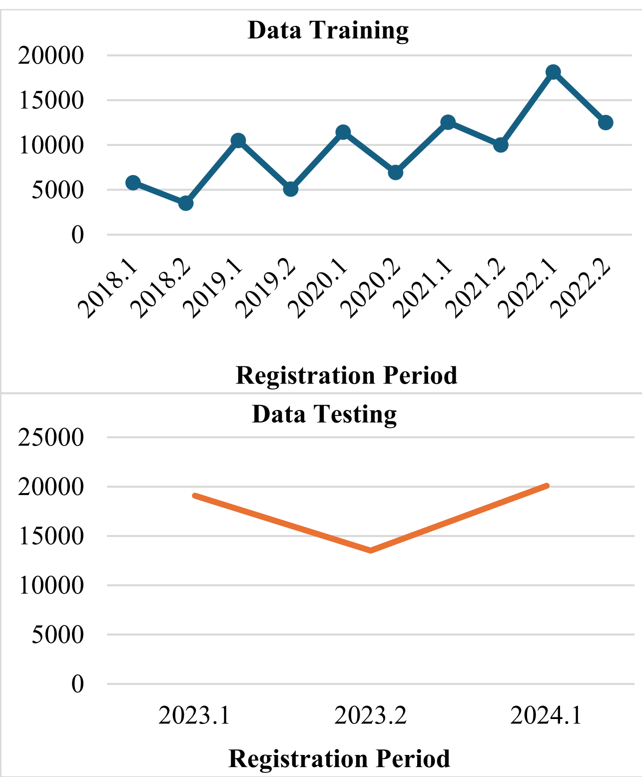

Fig. 4 Data training and data testing

In this study, the parameters used in the Auto-ARIMA forecasting model are as follows:

1. p (Autoregressive Order)
2. d (Differencing Order)
3. q (Moving Average Order)

These parameters were determined through a hypertuning process using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are commonly employed to identify the optimal ARIMA model for forecasting.

In Python, AIC is used to compare the goodness-of-fit among various combinations of p, d, and q values.

The model with the lowest AIC is selected as the best-fitting model. Based on the parameter selection process, the ARIMA (2,1,0) model with p = 2, d = 1, and q = 0 was found to provide the most accurate predictions against the testing data, achieving an AIC of 34.512 and a BIC of 30.591, as shown in the following results:

SARIMAX Results

```
==========================================================
                    =======================
Dep. Variable:            jumlah  No. Observations:          3
Model:            ARIMA(2, 1, 0)  Log Likelihood        -14.256
Date:           Wed, 11 Dec 2024  AIC                    34.512
Time:                   07:32:13  BIC                    30.591
Sample:                        0  HQIC                   26.312
                             - 3
Covariance Type:             opg
==========================================================
             coef     std err        z      P>|z|    [0.025    0.975]
----------------------------------------------------------
ar.L1     -0.0680     1.137     -0.060    0.952    -2.296    2.160
ar.L2      0.9189     1.143      0.804    0.422    -1.322    3.160
sigma2  2.518e+04  2.21e-06  1.14e+10   0.000   2.52e+04  2.52e+04
==========================================================
Ljung-Box (L1) (Q):          2.00  Jarque-Bera (JB):        0.33
Prob(Q):                     0.16  Prob(JB):                0.85
Heteroskedasticity (H):       nan  Skew:                    0.00
Prob(H) (two-sided):          nan  Kurtosis:                1.00
==========================================================
```

Based on the dataset of teaching material usage across several academic programs and using the same parameter settings, the forecasting results from these criteria generated using the best ARIMA model (2,1,0) are presented in the following visualizations:

1. Teaching Materials for EKMA4159 in 3 Programs Study

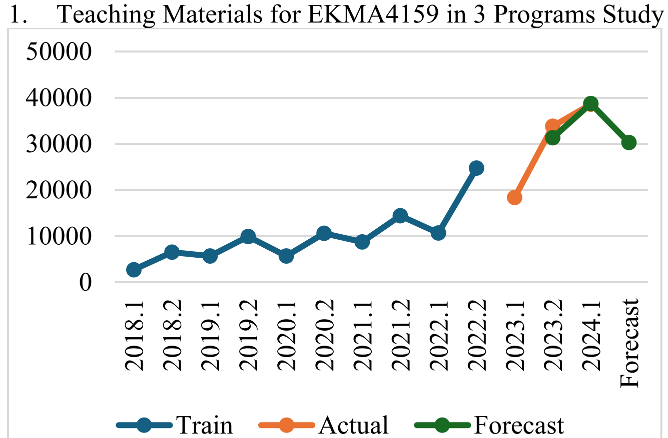

Fig. 5 Forecasting teaching material for EKMA4159

Based on the forecasting calculation for the teaching material EKMA4159, which is used in three study programs, and applying the best ARIMA model (2,1,0), the results showed a forecasting accuracy of 92.66% and a potential overestimation of teaching material needs by 0.37%.

2. Teaching Materials for ESPA4111 in 5 Programs Study

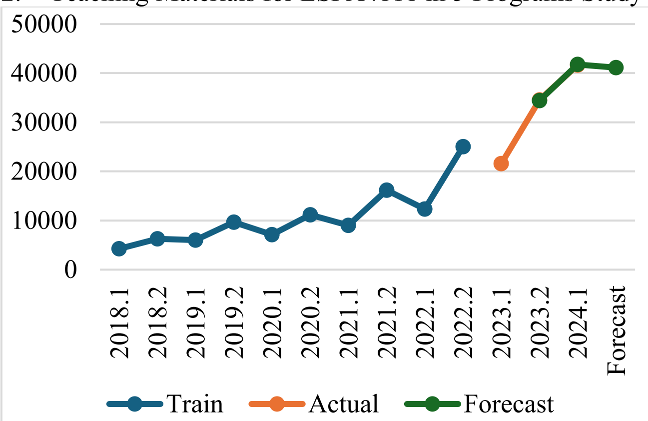

Fig. 6 Forecasting teaching material for ESPA4111

Based on the forecasting calculation for the teaching material ESPA4111, which is used in five study programs, and applying the best ARIMA model (2,1,0), the results showed a forecasting accuracy of 99.55% and a potential overestimation of teaching material needs by 0.19%.

3. Teaching Materials for EKMA4434 in 11 Programs Study

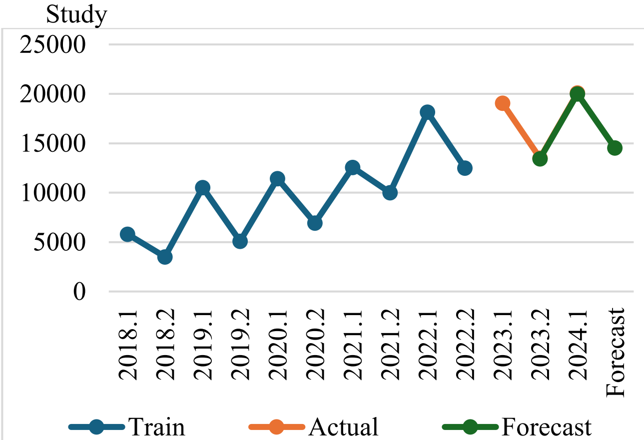

Fig. 7 Forecasting teaching material for EKMA4434

Based on the forecasting calculation for the teaching material EKMA4434, which is used in eleven study programs, and applying the best ARIMA model (2,1,0), the results showed a forecasting accuracy of 99.64% with no projected overestimation in teaching material needs, thereby eliminating the risk of excess inventory that could lead to material disposal due to expiration.

4. Teaching Materials for MKDU4109 in 13 Programs Study

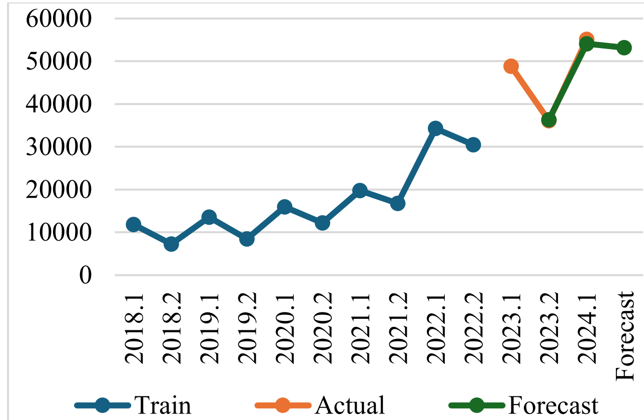

Fig. 8 Forecasting teaching material for MKDU4109

Based on the forecasting calculation for the teaching material MKDU4109, which is used in thirteen academic programs, and applying the best ARIMA model (2,1,0), the results showed a forecasting accuracy of 98.21% and a potential overestimation of teaching material needs by 0.50%

5. Teaching Materials for ISIP4112 in 15 Programs Study

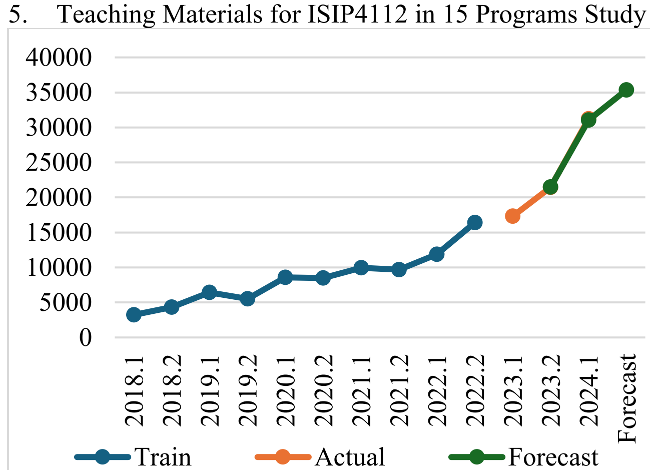

Fig. 9 Forecasting teaching material for ISIP4112

Based on the forecasting calculation for the teaching material ISIP4112, which is used in fifteen study programs, and applying the best ARIMA model (2,1,0), the results showed a forecasting accuracy of 99.54% and a potential overestimation of teaching material needs by 0.07%.A more detailed summary of the forecasting results for each criterion of teaching material usage across study programs is presented in the following table:

Table. 7 Forecasting result with ARIMA

Teaching Material	Number Of Program Study	Data Testing	Data Forecasting (Q1;Q2;Forecast)
EKMA4159	3	33.770;38.570	31.294;38.716;30.244
ESPA4111	5	34.571;41.685	34.418;41.768;41.168

EKMA4434	11	13.503;20.082	13.444;20.010;14.526
MKDU4109	13	36.125;55.101	36.306;54.115;53.163
ISIP4112	15	21.446;31.206	21.463;31.064;35.346

4.5. Evaluation

A ranking analysis was conducted after forecasting using the optimal ARIMA model, determined through hypertuning based on AIC and BIC criteria, which provided the values for parameters p, d, and q. This analysis was based on the accuracy level by comparing predicted values with actual data. The model that yielded the lowest evaluation scores in MAE, MSE, and RMSE tests was considered the best-performing model, as lower evaluation values indicate smaller prediction errors. The evaluation results are presented in the following table:

Table. 8 Forecasting accuracy test results

No	Teaching Material	Number Of Program Study	MODEL ARIMA	AIC	BIC	MAE	MSE	RMSE
1	EKMA4159	3	(2,1,0)	40,195	36,274	8.733.083	86.966.429.303	9.325.579
2	ESPA4111	5	(2,1,0)	36.590	32.670	6.850.611	72.297.411.954	8.502.788
3	EKMA4434	11	(2,1,0)	34.512	30.591	5.894.133	34.926.697.009	5.909.881
4	MKDU4109	13	(1,1,3)	44.162	37.628	10.824.745	161.609.496.043	12.712.572
5	ISIP4112	15	(2,1,0)	36.864	32.944	5.972.105	42.312.621.788	6.504.815

5. Conclusion and Future Work

5.1. Conclusion

Based on the forecasting of teaching material duplication needs using the Auto-ARIMA model, the best-performing model was identified as ARIMA (2,1,0), with parameters p = 2, d = 1, and q = 0, yielding an AIC of 34.512 and a BIC of 30.591. The teaching material EKMA4434, which is used in 11 academic programs, demonstrated the lowest error metrics across MAE, MSE, and RMSE, indicating the highest forecasting accuracy. This suggests that the forecasting performance improves when a teaching material is used across a greater number of programs, resulting in a more robust dataset. The RMSE value of 5,909.881 indicates that, on average, the forecast deviated from the actual figures by approximately 5,910 units. However, when applying ARIMA to a small dataset, the parameter estimations (p, d, q) tend to be less stable, which may hinder the model's ability to detect seasonal trends, long-term patterns, or autocorrelation structures.

As a result, forecasts may be less accurate and more vulnerable to noise or outliers. Nevertheless, the results of this forecasting study using the Auto-ARIMA model apply to Universitas Terbuka. With the addition of a larger and more diverse historical dataset, forecast accuracy is expected to improve further and support more efficient material procurement planning.

5.2. Future Work

Future research should focus on implementing causal forecasting methods, particularly neural network architectures. These models can incorporate distinguishing attributes and process larger, more complex datasets, potentially leading to higher forecast accuracy. Further comparative studies are needed to evaluate the performance of ARIMA models against neural network-based models. Such research will help identify which method is more effective and efficient in predicting teaching material requirements within distance education institutions like Universitas Terbuka.

Author Contributorship

Bagus Wicaksono: Conceptualization, Data Curation, Methodology, Resources, Investigation, Writing Original Draft, Project Administration, Visualization.
Tuga Mauritius: Conceptualization, Data Curation, Methodology, Resources, Writing Review and editing, Supervision, Project Administration, Validation.

Data Availability

Forecasting Teaching Materials Using the Autoregressive Integrated Moving Average (ARIMA) Method with data teaching materials expenditure Universitas Terbuka. The dataset used in this research is available at DOI: https://dx.doi.org/10.21227/gbhq-8y16

References

[1] George E. P. Box et al., *Time Series Analysis forecasting and Control*, Revised Edition, Holden Day, San Francisco, 1976. [Google Scholar]

[2] P. Sathish kumar et al., “Rainfall Forecasting using Machine Learning Techniques,” *Journal of Xidian University*, vol. 14, no. 6, pp. 2177-2189, 2020. [CrossRef] [Publisher Link]

[3] Sri Baginda Dalimunthe, Rosnani Ginting, and Sukaria Sinulingga, “The Implementation of Machine Learning in Demand Forecasting: A Review of Methods Used in Demand Forecasting with Machine Learning,” *Journal of Industrial Engineering Systems*, vol. 25, no. 1, pp. 41-49, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Deepali M. Kotambkar, and Pallavi M. Wankhede, "Hybrid LSTM/GRU-based Domain Adaptation Model for Correlation Analysis to Detect Glaucoma," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 1, pp. 168-175, 2023 [CrossRef] [Publisher Link]

[5] Adeline P. Dela Cruz et al., "Higher Education Institution (HEI) Enrollment Forecasting Using Data Mining Technique," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 2060-2064, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Alexiei Dingli, and Karl Sant Fournier, "Financial Time Series Forecasting - A Machine Learning Approach," *Machine Learning and Applications: An International Journal*, vol. 4, no. 1/2/3, pp. 11-27, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Josef Eiglsperger, Florian Haselbeck, and Dominik G. Grimm, "ForeTiS: A Comprehensive Time Series Forecasting Framework in Python," *Machine Learning with Applications*, vol. 12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hasbi Hassyddiqy, and Hasdiana, "Sales Forecasting Analysis Using ARIMA (Autoregressive Integrated Moving Average) Method at Huebee Indonesia," *Data Sciences Indonesia*, vol. 2, no. 2, pp. 92-100, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] G.O. Kyung, and I.F. Myung-Suck, "Evaluation of Short-Term Interval Models for Financial Time Series Forecasting," *SSRG International Journal of Industrial Engineering*, vol. 2, no. 3, pp. 5-8, 2015. [CrossRef] [Publisher Link]

[10] James J. Lee, and Misuk Lee, "Intelligent Structuration: Machine Learning Forecasting," *Issues in Information Systems*, vol. 23, no. 1, pp. 239-246, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xing Yee Leong et al., "Forecasting Elective Surgery Demand using ARIMA-Machine Learning Hybrid Model," *European Journal of Artificial Intelligence and Machine Learning*, vol. 2, no. 3, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Li Deng, and Dong Yu, "Deep Learning: Methods and Applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3-4, pp.197-387, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[13] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," *PLOS ONE*, vol. 13, no. 3, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Bohdan M. Pavlyshenko, "Machine-Learning Models for Sales Time Series Forecasting," *Data*, vol. 4, no. 1, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Gregory Piatetsky, "*CRISP- DM, Still the Top Methodology for Analytics, Data Mining, or Data Science Projects*," KDnuggets, 2014. [Google Scholar] [Publisher Link]

[16] Christoph Schröer, Felix Kruse, and Jorge Marx Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Models," *Procedia Computer Science*, vol. 181, pp. 526-534, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ashish Sedai et al., "Performance Analysis of Statistical, Machine Learning and Deep Learning Models in Long-Term Forecasting of Solar Power Production," *Forecasting*, vol. 5, no. 1, pp. 256-284, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Asha Sunki et al., "Time Series Forecasting of Stock Market Using ARIMA, LSTM and FB Prophet," *MATEC Web of Conferences, International Conference on Multidisciplinary Research and Sustainable Development (ICMED 2024)*, vol. 392, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Atwi Suparman, and Amin Zuhairi, "Distance Education for Sustainable Development: Lessons Learned from Indonesia," *UT-SEAMOLEC International Seminar "Open and Distance Learning for Sustainable Development*, Jakarta, 2004. [Google Scholar]

[20] Richard J. Tersine, *Principles of Inventory and Materials Management*, 3rd ed., Prentice-Hall International, New Jersey, USA, 1994. [Google Scholar] [Publisher Link]

[21] Dominik Wolff, and Fabian Echterling, "Stock Picking with Machine Learning," *Journal of Forecasting*, vol. 43, no. 1, pp. 81-102, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] G. Peter Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *Neurocomputing*, vol. 50, pp. 159-175, 2003. [CrossRef] [Google Scholar] [Publisher Link]