Comparative Study on Regression Algorithms for Predicting Price of Online Course: Udemy Case Study

Maximus Aurelius Wiranata[1], Theresia Ratih Dewi Saputri[2*]

[1,2]Informatics, School of Information Technology, Universitas Ciputra, Surabaya
e-mail: [1]maurelius@student.ciputra.ac.id, [2*] theresia.ratih@ciputra.ac.id

Submitted Date: 2023-06-07 Reviewed Date: 2023-06-08
Revised Date: 2023-06-27 Accepted Date: 2023-06-29

Abstract

Talent in the field of information technology is much needed. However, studying in the field of information technology requires a sizable fee. Online courses are a cost-effective option for learning. Online course sites like Udemy provide and sell hundreds of thousands of courses and have thousands of trusted instructors. With so many Udemy instructors, prices vary widely because the course pricing system is completely set by the teaching instructor. This means that the selling price of the course is not affected by the quality of the course, so not all courses are recommended to be purchased. To overcome this problem, a system is needed that can predict course prices so that it can advise instructors in determining selling prices. To compare the best algorithms used to create this system, three algorithms are used in this study: multiple linear regression, polynomial regression, and K-Nearest Neighbors Regression. The researcher uses 1200 data sample from web scraping results from the Udemy site, with one test for each algorithm. As a result, the K-Nearest Neighbors Regression got the best evaluation results with a root mean squared error value of 231659.49, a mean absolute percentage error of 0.43, and a coefficient of determination of 0.18.

Keywords: Comparative Study; Machine Learning; Price Prediction; Regression

1. Introduction

Digital development in Indonesia trails behind that of other nations such as India, China, and USA. According to the United Nations' (UN) 2020 e-Government Development Index (EDGI) survey, Indonesia ranks 88th out of 194 countries (United Nation, 2020). This classification is significantly lower than that of Singapore (11th) and Malaysia (47th). In Indonesia, the lack of human resources with digital expertise impedes digitalization. Unfortunately, Indonesia has few professionals with digital expertise. The market demand for digital workers exceeds the supply of digital workers (Fauzia, F., 2021).

Expanding education in the field of information technology is one of the ways to increase the number of digital workers. Unfortunately, the IT education is not avalaible in the entire segments of Indonesian Society. For many Indonesian’s citizens, high level education can be considered as a tertiary need due to its expensive cost (Fafirudin, T., 2021). Not every Indonesian has access to education in the sphere of information technology. The cost of tuition is an essential consideration for students with limited financial resources (Sumarmo, S., 2017). For instance, the University of Riau's Single Tuition Fee (*UKT*) is IDR 13,425,000 per semester (Kristen, U., 2017).

Online courses are a cost-effective option for education. There are numerous online course websites, such as Udemy, Dicoding, and Coursera, that offer various courses in the field of information technology. The reduced cost of online courses relative to tuition fees makes them accessible to those with a lower socioeconomic status. Udemy, one of the world's largest online course platforms, offers online courses for as little as IDR 99,000. However, there are still some amounts of people question whether the value of online courses commensurate with their quality.

Udemy grants instructors complete pricing autonomy. This pricing structure provides instructors with complete pricing autonomy. However, this pricing system has the disadvantage that not all courses' quality matches their selling price. A number of courses sold for more than one million rupiah have not been updated since 2013. Therefore, a system is required to determine the prices of online courses that are commensurate with their quality. In this study, we aim to conduct a comparative study on different machine learning algorithm to predict the course price using regression.

The compative study was hold by comparing three different regression algorithms such as Multiple Linear Regression, Polynomial Regression, and K-Nearest Neighbors Regression. Those algorithms were chosen due to their remarkable performance in handle continus data data prediction. Each algorithm was tested once with a data set consisting of 80% training data and 20% testing data. Three parameters are utilized in the testing evaluation: Root Mean Squared Error, Mean Absolute Percentage Error, and Coefficient of Determination. The best results from a comparison of the three algorithms will be attained through three tests to regress the price of Udemy courses.

2. Related Works

In the area of finance and economics, numerous studies and applications have conducted over the years that utilize machine learning algorithms to predict stock prices, real estate prices, commodity prices (Wiradinata, T, 2022; Behera, J, 2023). Traditionally, price prediction models were built using time series models such as ARIMA (Ariyo, A.A., 2014), but with the advancement of machine learning, there has been a shift towards using more complex algorithms. These models estimate future prices using historical data, market indicators, and economic factors. Due to its capacity to capture the relationships between independent variables and the objective variable, regression, a widely employed machine learning technique, has been widely employed in price prediction tasks.

The discipline of predictive analytics has been enhanced by machine learning, which offers compelling tools for analyzing and predicting various case. Machine learning algorithms show remarkable performance at handling complex and nonlinear relationships between predictors and the objective variable in price prediction (Chen,W, 2021). Machine learning, as a subset of artificial intelligence, has been particularly influential in financial analytics and price prediction due to its ability to process large volumes of data and find hidden patterns or trends. One of the task in machine learning is regression, used to predict dependent variable with continuous type. The regression models have been extensively used for price prediction tasks. The regression algorithms allow the estimation of the relationship between multiple independent variables and the dependent variable, allowing researchers to measure the impact of various factors on price changes. This strategy has proven effective at identifying the fundamental drivers of price fluctuations and producing accurate forecasts.

Due to their interpretability and robustness, regression models have long been a pillar in price prediction research (Madhuri, C.R, 2019). By showing the relationship between predictors and the objective variable as a linear or nonlinear equation, regression models provide a framework for comprehending the factors that influence price variations (Osborne, J.W., 2000). By understanding the factors, one can make informed predictions about how changes in one variable will affect another which can impoves the decision-making process. Various regression techniques, including multiple linear regression, polynomial regression, and support vector regression, have been employed by researchers to capture the complex relationships inherent in price data. These models allow the identification of significant predictors and the estimation of their influence on price fluctuations. In addition, regression models allow the assessment of the statistical significance of the predictors.

The evaluation of the overall model performance is measured using different metrics such as Mean Squared Error (MSE), the coefficient of determination (R-squared) and root mean squared error (RMSE) (Botchkarev, 2018). Researchers are able to make informed decisions based on the model's predictions by utilizing regression analysis to obtain insight into the primary determinants of price changes.

The advancements in machine learning and regression modeling have substantially impacted research into price prediction. Therefore, this study utilizes the machine learning algorithm to predict the course price using regression techniques. Regression models provide interpretability and robustness in quantifying the impact of predictors on price variations. By combining these methods has allowed us to develop accurate and insightful models for comprehending and predicting course price trends.

3. Research Method

In order to address the research problem, this study utilized the the Cross-Industry Standard Process Model for Data Mining (CRISP-DM). There are six stages in the CRISP-DM method, namely: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment (Rohman, M.A., 2022) as seen in Figure 1.

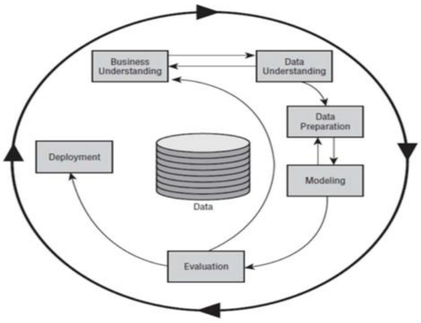


Figure 1. Adopted CRISP-DM Method by Rohman, M.A. (2022)

1. Business Understanding
   The first stage involves analyzing the attributes that can be measured to determine the price of online courses. These attributes include course ratings, category, number of students, the year when the course recently updated, and instructor's teaching history.

2. Data Understanding
   Data Understanding is a stage to comprehend and examine the data that will be used for model creation. This stage included the Explanatory Data Analysis (EDA) process that aims to understand the characteristic and pattern of each variable.

3. Data Preparation
   To generate a quality machine learning model, data preparation is necessary before processing. This process is carried out because the obtained data is often incomplete, inconsistent, and contains a lot of noise. Data preparation includes data cleaning, feature selection, and data transforms (Brownlee, J., 2020).

4. Modeling
   This stage involves creating a prediction model. In this research, three machine learning algorithms are used for the comparative purpose such as Multiple Linear Regression (linear regression), Polynomial Regression (non-linear regression), and K-Nearest Neighbors Regression.

a) Multiple Linear Regression
Regression is a process to predict the value of a target variable based on other features. Linear regression is one type of regression that models the features with the target variable as a linear equation. In Multiple Linear Regression, the model includes more than one explanatory variable (x1, x2,..., xp), resulting in a multivariate model (Tranmer, M., 2020). The following is the formula for Multiple Linear Regression:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

(Pane, 2021)

Where:
Y : Dependent Variable,
α : Constant,
β : Regression Coeficient,
X1, X2, …, Xn = Independent Variables

b) Polynomial Regression
Polynomial Regression is a regression technique that uses polynomial transformation before performing linear regression. The formula for Polynomial Regression is as follows (Atzzahra, H., 2021):

$$Y = a_o + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Where:
Y : Dependent Variable
a : Constant
$x_1, x_2, \ldots, x_n$: Independent Variables

c) K-Nearest Neighbors Regression
K-nearest neighbors (KNN) is a supervised learning algorithm that classifies objects based on the training data points that are closest to the object (Leidiyana, H., 2013). This algorithm can be used for both classification and regression tasks. The following is the formula for Euclidean Distance used in the KNN (Nishom, M., 2019):

ISSN: 2541-1004
e-ISSN: 2622-4615
10.32493/informatika.v8i2.30562

$$d(x,y) = |x-y| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Where:
d : distance
x : sample 1
y : sample 2
n : dimension

5. Evaluation

Evaluating the performance of the model is a stage to measure the quality of the used algorithm. The evaluation is measured using three metrics: Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R-squared).

a) RMSE

Root Mean Squared Error is the average of the squared differences between the actual values and the predicted values. A lower RMSE value indicates higher accuracy of the model's predictions. Researchers chose this metric to measure the magnitude of errors made by the model. The formula for RMSE is as follows (13):

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_1 - \widehat{y_1})^2}{n}}$$

Where:
y_i: observed value
$\widehat{y_i}$: predicted value
n: dimension

b) MAPE

MAPE Error is a method for measuring the absolute error of a model by calculating the deviation between actual values and predicted values (Krisma, A., 2019). The absolute errors are averaged and converted into a percentage, resulting in the Mean Absolute Percentage Error. Researchers use this metric to assess the model's forecasting capability.

A highly accurate model has a MAPE value below 10%, a good model has a MAPE value between 10% and 20%, a fair model has a MAPE value between 20% and 50%, while a poor model has a MAPE value above 50%. The formula for MAPE is as follows (Ginantra, N. L. W. S. R.2019):

$$MAPE = \frac{\sum \frac{|actual - forecast|}{actual} x100}{N}$$

Where:
Actual: target variable in data testing
Forecast: predicted target variable
N: number of data

c) R-Squared

The Coefficient of Determination is the proportion of the variance in the dependent variable that is explained by the regression model. It is used to measure the success of predicting the dependent variable based on the independent variables (Nagelkerke, N.J.D., 1991). With this metric, researchers can evaluate whether the selected features are suitable for the created model. The formula for the Coefficient of Determination is as follows (Anscombe, 1973):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

Where:
y_i: observed value
$\widehat{y_i}$: predicted value
n: dimension

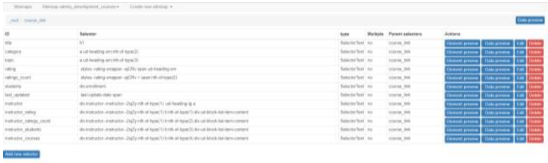
4. Result and Discussion

By applying the proposed methodology, this study is able to make a price prediction by comparing different regression algorithms. The result is discussed based on the step in the proposed methodology.

- Business Understanding

  This study uses data taken from the Udemy website using web scraping techniques. Web scraping was done using Google Chrome's Web Scraper extension. The total number of samples obtained during data collection process is 1200.

ISSN: 2541-1004
e-ISSN: 2622-4615
10.32493/informatika.v8i2.30562



ID	Selector	type	Multiple	Parent selectors
title	h1	SelectorText	no	course_link
category	[illegible]	SelectorText	no	course_link
topic	[illegible]	SelectorText	no	course_link
rating	[illegible]	SelectorText	no	course_link
ratings_count	[illegible]	SelectorText	no	course_link
students	div.enrollment	SelectorText	no	course_link
last_updated	[illegible]	SelectorText	no	course_link
instructor	[illegible]	SelectorText	no	course_link
instructor_rating	[illegible]	SelectorText	no	course_link
instructor_ratings_count	[illegible]	SelectorText	no	course_link
instructor_students	[illegible]	SelectorText	no	course_link
instructor_courses	[illegible]	SelectorText	no	course_link

Figure 2. Web Scrapper Extension

The procedure of web scraping is depicted in Figure 2. For scrapping the data, there are four column that are used as selector. Those columns that should be filled in include Id, selector, type, multiple, and parent selector. Id stands for the name of the selected selector, selector for the HTML element that was chosen for sampling, type for the selector's type, and multiple for the selector that was chosen one or many times. The Web Scraper Extension will begin scraping once all the selectors have been configured.

- Data Understanding
  Bar plots and box plots are used for data visualization in order to better understand the data they have used. Box plots are used to display the distribution of data, whilst bar plots are used to compare data for each variable.

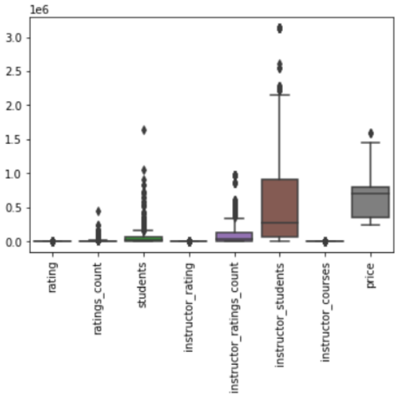


Figure 3. Value Range for all numeric variables

According to figure 3, the ratings_count, students,instructor_ratings_count, instructor_students, and price columns all include rather significant outliers. This shows that there is a large range of values in these columns and the distribution of the data.
Figure 4 shows that not every new course has a greater average cost. The courses with the highest average costs were last updated in 2013. It is thought that the most recent year a course was renewed has no real impact on course costs.

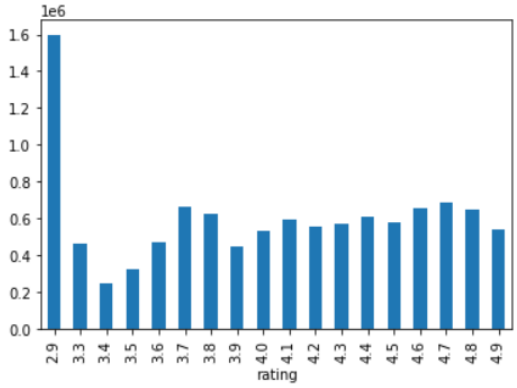


Figure 4. Average Course Price based on The Recent Update

Figure 5 demonstrates that the course with the highest average rating from its students is not the one with the highest average cost. Unexpectedly, courses with a rating of 2.9 have an extremely high average cost of IDR 1,600,000. In contrast, the cost of courses with a rating of 4.0 or higher ranges from IDR 400,000 to IDR 700,000. The data has an outlier rating of 2.9 according to this research, and data with other ratings have prices that aren't all that dissimilar from one another.

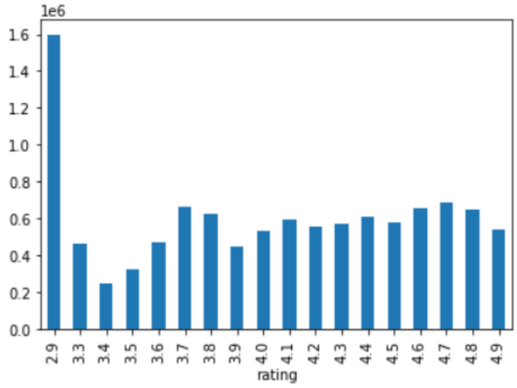


Figure 5. Average Course Price by Rating

The dependent variable and independent variables have a correlation between them that ranges from -0.029 to 0.33. The teacher rating count and category have the strongest correlations of 0.33 and -0.029, respectively. Because the correlation value is near to zero, it can be inferred that Category and course cost are unrelated. On the other hand, the instructor's overall grade has the biggest impact on the cost of the course. The average cost of the courses offered rises as more instructors receive positive feedback from students.

- Data Preparation

  Data preparation process was conducted in five stages.

  1. Data Cleaning

     Data cleaning is the first step in the data preparation process. In order to achieve better outcomes, data cleaning seeks to clean the data before the modeling process. Data cleaning is done in five stages, namely:

     a. Removing extra default columns from the Web Scraper addon The columns are course_link, course_link-href, web-scraper-order, and web-scraper-start-url.
     b. Students, instructor_rating, instructor_ratings_count, instructor_students, and instructor_courses should all be removed from the ratings_count column.
     c. Reduce the number of categories and delete "Last Updated" by removing the month from the Last Updated column. The words "Last Updated" appear in the initial data, and the format is month/year. Therefore the month information and "Last Updated" description are removed, leaving only the year numbers in the data.
     d. Changing columns with numeric data of type object to integer.
     e. Remove data that has no price information.

     There are a total of 1008 data samples with 10 features once the data has been cleaned, including:

     a. Category: the course category.
     b. Rating: course ratings from students who have purchased it.
     c. Rating count: the number of students who give ratings to the course.
     d. Students: the number of students who purchased the course.
     e. Last updated: the last year the course was renewed.
     f. Instructor rating: instructor ratings of students who have purchased courses from that instructor.
     g. Instructor rating count: the number of students who give ratings to the instructor.
     h. Instructor students: the number of the instructor's students.
     i. Instructor courses: the number of courses provided by the instructor.
     j. Price: the course price.

  2. Label Encoding

     Label encoding is used to assign a number label to the category column after obtaining clean data. As categories of courses, categorical data values require a number designation to be processed by a computer. An illustration of a label would be the Software Testing category, which would alter all of the category values to 8.
  3. Determination of dependent and independent data

     The price column from this study serves as the independent data, and all other columns serve as the dependent data.
  4. Data Normalization

     Data normalization is a necessary step in the standardization of data information. The MinMaxScaler technique converts all dependent variables into a range between 0 and 1. The Sklearn package MinMaxScaler, which is part of Sklearn preprocessing, provides assistance in the normalizing process.
  5. Splitting Training and Testing Data

     In this study, 80% of the data is used for training and 20% for testing. To distinguish between training and testing data, the Sklearn library train_test_split is used, which is part of the Sklearn model selection.

- Modeling

  Three algorithms—multiple linear regression, polynomial regression, and K-nearest neighbors regression—are used to build the model. The Sklearn library is used to create the three models.

- Evaluation

  The performance of each model will be evaluated using the calculations of the root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination(R^2).

Table 1. Evaluation of Test Results

Model	RMSE	MAPE	R^2
Multiple Linear Regression	242274.5	0.47	0.11
Polynomial Regression	289578.4	0.45	-0.26
K-Nearest Neighbors Regression	231659.49	0.43	1.87

The three experiments in Table 1 were carried out in this study, one for each algorithm. Multiple linear regression, the first algorithm put to the test, received a score of 242274.5044 on the root mean squared error measure, indicating that the model's error is rather high. The poor Coefficient of Determination score of 0.111449101, which indicates that the features utilized are inappropriate for usage in this model, supports this. The model's capacity to make predictions is characterized as practicable when the mean absolute percentage error value is between 0.20 and 0.50, which results in a value of 0.471970988.

When compared to multiple linear regression, which receives a root mean squared error score of 289578.4107, the second approach, polynomial regression, performs worse. In other words, the model's prediction mistakes are worse. In contrast, even if the mean absolute percentage error score, which is 0.459074802, is still in the right group, it receives a higher rating. The Coefficient of Determination for Polynomial Regression is -0.26940219. If the coefficient of determination is negative, the model's prediction is worse than a function that consistently forecasts the average of the data.

The K-nearest neighbors Regression was the final algorithm to be tested; it received the lowest root mean squared error score of 231659.4916, indicating that it has the lowest error rate. The mean absolute percentage error is also less, but at 0.431949334, it still falls within the acceptable range. The best Coefficient of Determination value for any algorithm is 0.187605301, which shows that some attributes are better suited for use in this algorithm.

Overall, evaluation findings for multiple linear regression, polynomial regression, and K-nearest neighbors regression were negative. The K-nearest neighbors regression, which has the lowest root mean squared error and mean absolute percentage error of the three models, is the most effective model. The K-nearest neighbors regression has the lowest error value and the best relations between columns of the three scores, it can be concluded.

All models are presumed to give subpar estimates since they have poor correlations with the price variable, which are below 0.5 for each variable. Since Udemy permits teachers to choose their own class fees, none of the test's factors have an impact on pricing. According to the evaluation's findings, the developed model can forecast the cost of Udemy courses. The evaluation's findings indicate that, to increase the precision of this regression model, modifications are required.

5. Conclusion

The results are not favorable because there is little association between the features. The K-nearest neighbors Regression algorithm with a value of k = 8 with a root mean squared error of 231659.49, a mean absolute percentage error of 0.43, and a coefficient of determination of 0.18 is the best algorithm according to the comparison's findings.

6. Future Work

The root mean squared error, mean absolute percentage error, and coefficient of determination have unfavorable values, indicating that the regression method needs to be improved. Future studies should consider other elements that affect course pricing, such as the course's content, the number of articles and instructional videos, and the length of time needed to complete it.

Acknoledgment

We acknowledge the funding from the Institute of Research and Community Development, Universitas Ciputra, Surabaya under the International Research Internal Funding Scheme (contract number: 003/UC-LPPM/DIP/SP3H/VIII/2022).

References

Anscombe, F. J. (1973). Graphs in statistical analysis. *The american statistician*, 27(1), 17-21.

Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. *In 2014 UKSim-AMSS 16th international* *conference on computer modelling and simulation* (pp. 106-112). IEEE.

Atzzahra, H. (2021). *ANALISIS SENSITIVITAS PENGARUH KEBIJAKAN PEMERINTAH DAN PENERAPAN POLYNOMIAL REGRESSION PADA MODEL TRANSMISI COVID-19* (Doctoral dissertation, Institut Teknologi Kalimantan).

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, 120, 105843.

Botchkarev, A. (2018). Evaluating performance of regression machine learning models using multiple error metrics in azure machine learning studio. *Available at SSRN 3177507*.

Brownlee, J. (2020). *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery.

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, 100, 106943.

Fafirudin, T., Fitriani, F., & Wulandari, A. (2021). Minat Mahasiswa Melanjutkan Kuliah: Intensitas Promosi, Kepercayaan dan Biaya Kuliah. *Jurnal Pengembangan Wiraswasta*, 23(3), 185-192.

Fauzia, F., Virantika, A., & Firmansyah, G. (2021). Langkah langkah Strategis Pemenuhan Kebutuhan SDM Talenta Digital di Lingkungan Pemerintahan Indonesia. *Proceeding KONIK (Konferensi Nasional Ilmu Komputer)*, 5, 39-46.

Ginantra, N. L. W. S. R., & Anandita, I. B. G. (2019). Penerapan Metode Single Exponential Smoothing Dalam Peramalan Penjualan Barang. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 3(2), 433-441.

Hastomo, W., Karno, A. S. B., Kalbuana, N., Nisfiani, E., & Lussiana, E. T. P. (2021). Optimasi Deep Learning untuk Prediksi Saham di Masa Pandemi Covid-19. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 7(2), 133-140.

Krisma, A., Azhari, M., & Widagdo, P. P. (2019, September). Perbandingan metode double exponential smoothing dan triple exponential smoothing dalam parameter tingkat error mean absolute percentage error (mape) dan means absolute deviation (mad). *In Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi* (Vol. 4, No. 2).

Kristen, U., Wacana, S., Tua, N., & Gaol, L. (2017). Magister Manajemen Pendidikan FKIP Teori dan Implementasi Gaya Kepemimpinan Kepala Sekolah. *Ejournal. Uksw. Edu*.

Leidiyana, H. (2013). Penerapan Algoritma KNN untuk Penentuan Resiko kredit Kepemilikan Kendaraan Bermotor. Jurnal Penelitian Ilmu Komputer Sistem Embedded dan Logic, 1(1), 65-76.

Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019, March). House price prediction using regression techniques: A comparative study. *In 2019 International conference on smart structures and systems (ICSSS)* (pp. 1-5). IEEE.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691-692.

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika*, 4(01), 20-24.

Osborne, J. W. (2000). Prediction in multiple regression. *Practical Assessment, Research, and Evaluation*, 7(1), 2.

Pane, S. F., Poetra, C. K., & Fatonah, R. N. S. (2021). Analisa Profit Dan Loss Pada Sistem Manajemen Aset Dengan Menggunakan Algoritma Multiple Linear Regression. Jurnal SITECH: Sistem Informasi dan Teknologi, 4(1), 1-6.

Rohman, M. A., & Harini, S. (2022). Komparasi Algoritma Naïve Bayes dan k-Nearest Neighbor Pada Klasifikasi Kontribusi Tokoh Politik. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS: Journal of Information System*, 7(1), 21-30.

Sumarno, S., Gimin, G., & Nas, S. (2017). Dampak Biaya Kuliah Tunggal Terhadap Kualitas Layanan Pendidikan. *Kelola: Jurnal Manajemen Pendidikan*, 4(2), 184-194.

Tranmer, M., & Elliot, M. (2008). Multiple linear regression. The Cathie Marsh Centre for Census and Survey Research (CCSR), 5(5), 1-5.

United Nations. (2020). UN&nbsp; E-Government&nbsp; Survey&nbsp; 2020. https://publicadministration.un.org.

Wiradinata, T., Graciella, F., Tanamal, R., Soekamto, Y. S., & Saputri, T. R. D. (2022). Post-Pandemic Analysis of House Price Prediction in Surabaya: A Machine Learning Approach. *Journal of Southwest Jiaotong University*, 57(5).