Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan
https://jurnal.stiq-amuntai.ac.id/index.php/al-qalam
P-ISSN: 1907-4174; E-ISSN: 2621-0681
DOI : 10.35931/aq.v17i3.2222

STOCK PRICE PREDICTION OF PT. KIMIA FARMA, TBK USING BAYESIAN RIDGE ALGORITHM

Aldy Mercyano Iqbal
Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia, 12260
2111601189@student.budiluhur.ac.id
Indra Tresna Setiadi
Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia, 12260
2111601130@student.budiluhur.ac.id
Aditya Dhian Pratama
Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia, 12260
2111601858@student.budiluhur.ac.id
Imelda Imelda
Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia, 12260
imelda@budiluhur.ac.id

Abstract

Stocks are proof of ownership of capital/funds in the company, a paper that clearly states the nominal value, and company name and is followed by an explanation of the rights and obligations to each shareholder, enough shares to be sold. Many companies from various sectors market their shares, one of which is PT. Kimia Farma, Tbk. This study aims to determine the performance of the Bayesian Ridge Regression algorithm to predict the closing price of PT Kimia Farma, Tbk's stock price. The data for these stocks was collected through the website id.investing.com with a data period of 2015 - 2022. Based on the test results, the Bayesian Ridge Regression algorithm can be used to predict stock prices with a good value where the R2 value is 0.9968 and RMSE value of 49.199 when modelling, R2 value is 0.9972 and RSME value is 45.400 when testing with new/testing data.
Keywords: Stock Prediction, PT Kimia Farma, Tbk., Bayesian Ridge Regression

Abstrak

Saham adalah bukti kepemilikan modal/dana dalam perusahaan, kertas yang dengan jelas mencantumkan nilai nominal, nama perusahaan dan diikuti penjelasan hak dan kewajiban kepada masing-masing pemegang saham, saham yang cukup untuk dijual. Banyak perusahaan dari berbagai sektor yang memasarkan sahamnya, salah satunya adalah PT. Kimia Farma, Tbk. Tujuan dari penelitian ini adalah untuk mengetahui performansi dari algoritma Bayesian Ridge Regression untuk memprediksi harga Penutupan/Terakhir harga saham PT Kimia Farma, Tbk. Adapun data untuk saham tersebut, dikumpulkan melalui situs id.investing.com dengan periode data 2015 - 2022. Berdasarkan hasil pengujian, algoritma Bayesian Ridge Regression dapat digunakan untuk memprediksi harga saham dengan nilai yang baik dimana nilai R2 sebesar 0,9968 dan nilai RMSE sebesar 49,199 saat dilakukan pemodelan, nilai R2 sebesar 0,9972 dan nilai RSME sebesar 45,400 saat dilakukan pengujian dengan baru/data pengujian.
Kata Kunci: Prediksi Saham, PT Kimia Farma, Tbk., Bayesian Ridge Regression

INTRODUCTION

Stock is proof of ownership of capital/funds in company, paper that clearly stated the nominal value, company name and followed by the rights and obligations explained to each shareholder, sufficient stock to sell.[1] Other than that, stock is proof of capital participation in a company. By buying company stocks, it means that you invest capital / funds that the management will later use to finance the company's operational activities.[2] Stock are instruments of capital participation so that shares are basically endowment funds, meaning that the participation will continue if the company is still standing.[3]

Many companies from various sectors market their shares, one of which is PT. Kimia Farma, Tbk. PT Kimia Farma, Tbk is a state-controlled pharmaceutical company in Indonesia. PT Kimia Farma, Tbk produces, markets, and distributes chemicals, medicines, biologics and other materials needed to produce pharmaceutical preparations, contraceptives, cosmetics, traditional medicines, medical devices, food/beverage products and other products including in plantation and mining sector.[4] Moreover, PT. Kimia Farma, Tbk is one of the 10 largest drug manufacturers which produces 334 types of drugs of which 60% are generic drugs and 40 percent are branded drugs.[5]

Stock price prediction are interesting to discuss, not only for economists but also analyst. Various technique has been made for predicted the stock price, one of them is machine learning. Machine learning is a branch of artificial intelligence. Using computing, we design system that can learn from data in a manner of being trained. The systems might learn and improve with experience and, with time, refine a model that can be used to predict outcomes of questions based on the previous learning.[6] Besides that, the development of machine learning is currently very developed and currently there is a library that makes it easy to develop machine learning models and one of them is Pycaret.

PyCaret is one of the easiest-to-use Python packages for AutoML, although it is still relatively new.[7] Automated ML, abbreviated and popularly known as AutoML, is a process of applying automation to the ML life cycle with the aim to automate the repetitive task of it. This will provide an edge to the technology by not only democratizing it and making it accessible to all but will have various advantage like reduction of the model run time, a well-tuned model, various

[1] S. Handini, *Buku Ajar : Manajemen Keuangan* (Surabaya: Scopindo Media Pustaka, 2020).
[2] A. P. Tambunan, *Menilai Harga Wajar Saham* (Jakarta: Elex Media Komputindo, 2008).
[3] H. M. Fakhruddin, *Go Public* (Jakarta: Elex Media Komputindo, 2013).
[4] B. Setianto, *Saham-Saham Industri Consumer Goods Di BEI per Laporan Keuangan Q4 2018* (BSK Capital, 2019).
[5] James J. Spillane, *Ekonomi Farmasi* (Jakarta: Grasindo, n.d.).
[6] J. Bell, *Machine Learning - Hands-On for Developers and Technical Professionals* (New Jersey: Wiley, 2020).
[7] N. George, *Practical Data Science with Python - Learn Tools and Techniques from Hands-on Examples to Extract Insights from Data* (Birmingham: Packt Publishing, 2021).

evaluation metrics to judge the model performance, etc.[8] The developer who created PyCaret is Moez Ali. PyCaret basically a wrapper around several machine learning libraries such as Scikit-Learn, XGBoost, Microsoft LightGBM, spaCy and more. Pycaret is a "low-code" library, and it is best known for its ease of use and efficiency.[9] The purpose of PyCaret is to implement the training, comparison, evaluation, optimization, interpretation, and distribution of machine learning modes with a few lines code, and consequently to automate and accelerate the cycle of experiments to improve productivity.[10] The library also provides an option to explain machine learning models with game-theoritic algorithm such as SHAP (Shapely Additive Explanation), also with single function call.[11]

Previously, research has been conducted on stock price predictions and majority used LSTM method. The result of previous research is presented below.

No.	Title	Author	Method/Algorithm	Result
1.	Stock Price Prediction Using the Long Short Term Memory Method In Pandemi Conditions	Endy Gigih Pratama, Intan Yuniar Purbasari, Wahyu S. J. Saputra	LSTM	The best test is a model with 3 hidden layer parameters of 50,100,150 neurons, a learning rate of 0.001, with an error value calculation using RMSE of 0.0015 and calculations using MAPE of 20%.
2.	Mining Stock Price Forecasting on the Indonesia Stock Exchange (IDX) Using Long Short Term Memory (LSTM)	Roby Julian, Muhammad Rizky Pribadi	LSTM	By using the Long Short Term (LSTM) this study produces a fairly good RMSE value with an increase in the RMSE value based on the addition of the number of epoch variations. The optimal epoch variation was obtained with the number of epochs of 200. Meanwhile, the optimal RMSE value produced by the Long Short Term Memory (LSTM) method was generated by TINS issuers with an RMSE of 31.71.
3.	Prediction of Stock	Ardiyan	LSTM	By using hidden layer

[8] A. Abraham and A. K. Tyagi, *Data Science for Genomics* (Elsevier Science, 2022).

[9] P. Xiao, *Artificial Intelligence Programming with Python* (New Jersey: Wiley, 2022).

[10] M. Cannataro et al., *Artificial Intelligence in Bioinformatics - From Omics Analysis to Deep Learning and Network Mining* (Elsevier Science, 2022).

[11] D. Radecic, *Machine Learning Automation with TPOT - Build, Validate, and Deploy Fully Automated Machine Learning Models with Python* (Birmingham: Packt Publishing, 2021).

	Price Movements in the Pharmaceutical Sector Using the Long Short-Term Memory Algorithm	Agusta, Iin Ernawati, Anita Muliawati		parameters, units as well as epoch and batch size variations which produce stock price prediction results with an average RMSE value of 27,310.
4.	Cyclical Stock Price Prediction in Indonesia Using the LSTM and SVM Methods	Gabriel Adisurya Harsono, Alexander Setiawan, Hans Juwiantho	LSTM & SVM	From this research found 1 best metode for multivariate prediction and one for univariate. for multivariate the best parameter for the prediction is LSTM where the RMSE value is 63.67 training and 74.82 for testing. For univariate prediction the best metode is SVR where the RMSE value is 58.04 for training and 75.29 for testing.
5.	Comparative Analysis of Model Prediction in Stock Price Prediction Using Linear Regression, Random Forest Regression and Multilayer Perceptron Methods	Evita Fitri, Dwiza Riana	Linear Regression, Random Forest Regression & Multilayer Preceptron	The prediction model with LR is able to produce a fairly low error prediction value with the lowest RMSE score of 0.010 and the highest RMSE of 0.012, the lowest MAPE of 1.2% and the highest of 1.9%, the lowest MAE of 0.006 and the highest. of 0.009, and the highest R2 value was 99.8% and the lowest was 99.6%. It can be concluded that in this study, the Linear Regression prediction model is able to predict historical data on stock prices better than the RFR and MLP models.
6.	Comparison of Long Short-Term Memory Algorithm with SVR on Stock Price Prediction in Indonesia	Adhib Arfan, Lussiana ETP	LSTM & SVR	Based on the test results, LTSM is able to predict stock prices in 2017-2019 in good performance and a relatively little error rates. The test used the Support Vector Regression (SVR) method, LSTM has a better loss value than the SVR algorithm. The data range in the LSTM affects the training time used. The bigger data range is, the

				longer training time is used. The data range on the SVR affects the loss value. The bigger data range is, the bigger the loss value is generated. Thus, it can be concluded that LSTM is able to overcome long-term dependency and able to predict stock prices with accurate result.

Based on the previous research results above, we will conduct research to find other algorithms/methods that can be used in predicting stock prices. And in this study will discuss the bayesian ridge regression algorithm to determine the closing price of shares and also evaluate the performance of the algorithm using the Pycaret library.

MACHINE LEARNING MODEL PIPELINE

Share price data for PT. Kimia Farma, Tbk is collected from id.investing.com site with data ranging from 2015 to 2022. The data obtained regarding stock prices provided by the site are the date of stock price, closing price, opening price, highest price, lowest price, volume, and the percentage change in the closing price of shares. after the data is obtained, then the next is pre-processing. In pre-processing, the data that has been retrieved will be divided into 80:20 where 80% of the data will later be used in the modeling process and 20% will be used for testing new data (unseen data). after the pre-processing is complete, then do the initial configuration of the PyCaret library which will later be used in making machine learning models using the bayesian ridge regression. In this initial configuration, the 80% data that we divided before, will be split again into 70:30, where 70% is training data and 30% is testing data.

RESEARCH METHODS

Data Collection

Table 1. Stock Price of PT Kimia Farma, Tbk

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%
24/11/2022	1.295	1.29	1.3	1.28	406,70K	0,00%
23/11/2022	1.295	1.335	1.365	1.29	1,52M	-2,63%
22/11/2022	1.33	1.4	1.4	1.315	2,02M	-5,00%
21/11/2022	1.4	1.42	1.445	1.38	1,32M	-1,06%
18/11/2022	1.415	1.38	1.45	1.355	2,09M	2,54%
17/11/2022	1.38	1.41	1.425	1.335	1,69M	-1,78%
16/11/2022	1.405	1.385	1.46	1.385	3,45M	1,44%
15/11/2022	1.385	1.51	1.77	1.385	32,65M	-6,73%
14/11/2022	1.485	1.19	1.485	1.19	13,46M	24,79%
11/11/2022	1.19	1.185	1.2	1.18	252,60K	0,42%

"*Tanggal*" is the date of stock were published in the market on the period. "*Terakhir*" is the closing price of the stock. "*Pembukaan*" is the stock price data traded for the first time that day. "*Tertinggi*" and "*Terendah*" are the day's highest and lowest stock prices. "*Vol.*" is an amount of an asset or stock, that changes hands over a period of time, and "*Perubahan%*" is the change of closed price of the stock, comparing with data on previous day. In this study, the focus is on to predict the closing price ("*Terakhir*") as a dependent variable and value of "*Pembukaan*", "*Tertinggi*" and "*Terendah*" as the independent variable.

Pre-Processing

In the pre-processing stage, the first thing to do is to select the columns to be used in building a machine learning model, where in this case the column used is "*Pembukaan*", "*Tertinggi*" and "*Terendah*" as the indpendent variable and column "*Terakhir*" as the dependent variable/target. The unused columns then dropped from the dataframe. After that, the data is divided into 80:20, where 80% of the data is used for building model and 20% for testing data after creating the model (unseen data/new data).

	Terakhir	Pembukaan	Tertinggi	Terendah
0	1315	1315	1315	1305
1	1085	1120	1120	1085
2	985	1000	1000	985
3	730	705	765	700
4	640	635	660	620

Figure 1. Selected Columns using for Modeling

Table 2. Spliting Data for Modeling and Testing After Modeling

No	Remarks	Total Data
1.	Data for Modeling	1536
2.	Data for testing after modeling	384

RESULTS AND DISCUSSION

Configuration and Building Model

One of the most useful Bayesian ridge techniques is Bayesian ridge regression which estimates a probabilistic model of the regression problem to predict continuous values. Bayesian ridge regression does not need to have any extra prior knowledge about the dataset.[12] Bayesian regression makes it possible to deal with insufficient data or poorly distributed data by formulating linear regression using probability distributors rather than point estimates. Mathematically, to get a full probabilistic model, the output is assumed to be Gaussian distribution around X_w as follows:

$$p(y|X, w, \alpha) = N(y|X_w, \alpha)$$

Where p is probability, y is output, X is input data, w is parameter or weights, α is hyperparameter related to variance, N is normal distribution and X_w is the mean. Below is the prior for the coefficient w is given by spherical Gaussian as follows.

$$p(w|\lambda) = N\left(w|0, \lambda^{-1} I_p\right)$$

Where p is probability, w is parameter or weights, λ is hyperparameter related to variance, N is normal distribution and 0 is zero mean.

Before creating a model using pycaret, the first step that needs to be done is to configure pycaret. The configuration steps are as follows.

```
import pandas as pd
from pycaret.regression import *
```

Figure 2. Import Library

The first line we import the pandas library which will be used to load PT Kimia Farma, Tbk stock data and the second line we call all the regression methods contained in the Pycaret library.

```
In [9]: #buat experimen dengan pycaret modul regresi
        exp_reg = setup(data = data, target='Terakhir', session_id=100)
```

Figure 3. Setup PyCaret

[12] S. C. Haw and K. S. Muthu, *Proceedings of the International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022)* (Atlantis Press, 2022).

After importing library, do the setup for model creation with call setup function. Setup function initializes the experiment in PyCaret and creates the transformation pipeline based on all the parameters passed in the function. Setup function must be called before executing any other function. It takes two mandatory parameters: data and target. All the other parameters are optional.[13] After that, we will compare all models that contained in the PyCaret library. Using setup function and comparing the models in the PyCaret library which shown in the image below.

```
In [10]: compare_models()
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lr	Linear Regression	29.9143	2509.2494	49.2324	0.9968	0.0212	0.0141	0.9750
lasso	Lasso Regression	29.9175	2509.2328	49.2328	0.9968	0.0212	0.0141	0.3190
ridge	Ridge Regression	29.9145	2509.2637	49.2326	0.9968	0.0212	0.0141	0.0080
en	Elastic Net	29.9157	2509.2410	49.2326	0.9968	0.0212	0.0141	0.0080
lar	Least Angle Regression	29.9142	2509.2374	49.2323	0.9968	0.0212	0.0141	0.0080
br	Bayesian Ridge	29.9102	2509.1816	49.2311	0.9968	0.0212	0.0141	0.0070
huber	Huber Regressor	28.9221	2752.6261	51.0518	0.9966	0.0215	0.0135	0.0100
omp	Orthogonal Matching Pursuit	37.1331	3396.8122	57.4917	0.9956	0.0257	0.0176	0.0070
llar	Lasso Least Angle Regression	41.4030	3485.2943	58.0934	0.9956	0.0313	0.0228	0.0070
et	Extra Trees Regressor	31.2514	3707.1732	58.5584	0.9955	0.0238	0.0145	0.1030
gbr	Gradient Boosting Regressor	33.9710	3741.9014	59.7709	0.9954	0.0238	0.0155	0.0190
rf	Random Forest Regressor	31.1216	3898.3589	59.7794	0.9952	0.0233	0.0142	0.1160
par	Passive Aggressive Regressor	42.5674	4455.7507	61.4471	0.9942	0.0260	0.0196	0.0080
knn	K Neighbors Regressor	32.5563	5286.1711	66.0794	0.9939	0.0240	0.0146	0.0210
dt	Decision Tree Regressor	37.6337	6132.0583	76.2577	0.9924	0.0287	0.0171	0.0090
ada	AdaBoost Regressor	70.8627	9889.8298	98.6485	0.9876	0.0577	0.0397	0.0230
lightgbm	Light Gradient Boosting Machine	44.9498	18996.1403	119.9804	0.9785	0.0340	0.0182	0.0810
dummy	Dummy Regressor	752.1409	790293.8723	886.9468	-0.0059	0.4668	0.4736	0.0070

Figure 4. Model Comparation with PyCaret

On the picture above, there was detected 6 models which have same R2 (R Square) score (the default metrics and sorted from highest to lowest value). In this research, the Bayesian Ridge model will be used which has the best metric result comparing other models which have same R2 value (have lowest value of RMSE). In Addition, the formulas of R2, RMSE were presented below.

$$R^2 = 1 - \frac{RSS}{TSS}$$

Where R^2 is coefficient of determination, RSS is sum of squares of residuals and TSS is total sum of squares.

$$RMSE = \left(\frac{\sum(y_i - \hat{y}_i)}{n}\right)^{1/2}$$

[13] M. Ali, "PyCaret: An Open Source, Low-Code Machine Learning Library in Python," PYCARET, 2020, https://www.pycaret.org.

Where RMSE is root mean square error value, y is observed value, ŷ is predicted value, i is the order of data and n is total data.

After that, we create the model that using Bayesian ridge regression with call function named “create_model”. In that function, we passed the parameter for creating the model. For the parameter value can be seen in figure 3 and in this study Bayesian ridge regression will be used with the parameter value “br”.

```
In [11]: br_reg = create_model('br')
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	30.1102	1929.9064	43.9307	0.9976	0.0204	0.0142
1	29.4095	2962.7976	54.4316	0.9963	0.0199	0.0136
2	33.5930	4201.2831	64.8173	0.9958	0.0208	0.0136
3	33.6326	3403.4871	58.3394	0.9953	0.0265	0.0157
4	24.6019	1252.9097	35.3965	0.9983	0.0187	0.0126
5	33.3954	2827.1075	53.1706	0.9959	0.0229	0.0155
6	30.2502	2460.5159	49.6036	0.9968	0.0221	0.0149
7	24.8783	1414.4858	37.6097	0.9980	0.0171	0.0121
8	26.8647	1566.3042	39.5766	0.9976	0.0182	0.0133
9	32.3664	3073.0183	55.4348	0.9968	0.0257	0.0152
Mean	29.9102	2509.1816	49.2311	0.9968	0.0212	0.0141
Std	3.2981	910.5641	9.2458	0.0010	0.0029	0.0012

Figure 5. Creating Bayesian Ridge Regression Model

The following details the default parameters used in the Bayesian ridge regression that has been created.

```
BayesianRidge(alpha_1=1e-06, alpha_2=1e-06, alpha_init=None,
              compute_score=False, copy_X=True, fit_intercept=True,
              lambda_1=1e-06, lambda_2=1e-06, lambda_init=None, n_iter=300,
              normalize=False, tol=0.001, verbose=False)
```

Figure 6. Default Parameters on Bayesian Ridge Regression Model

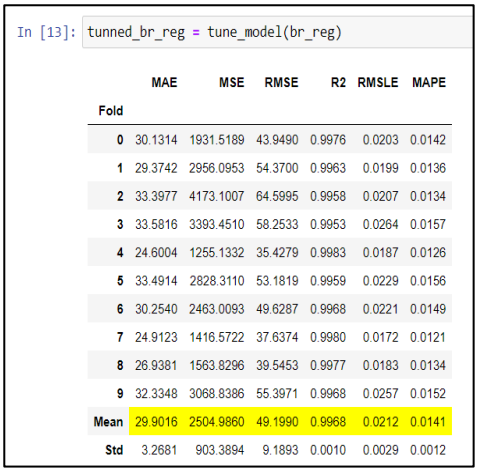
In addition, the following is an explanation for each parameter that used on Bayesian ridge regression.

Table 3. Description Parameter Bayesian Ridge Regression

Parameter	Description
alpha_1	Shape parameter for the Gamma distribution prior over the alpha parameter.
alpha_2	inverse scale parameter (rate parameter) for the Gamma distribution prior over the alpha parameter.
alpha_init	Initial value for alpha (precision of the noise)
compute_score	If True, compute the log marginal likelihood at each iteration of the

	optimization.
copy_X	If True, X will be copied; else, it may be overwritten.
fit_intercept	The intercept is not treated as a probabilistic parameter and thus has no associated variance. If set to False, no intercept will be used in calculations
lambda_1	Shape parameter for the Gamma distribution prior over the lambda parameter.
lamba_2	shape parameter for the Gamma distribution prior over the lambda parameter.
lambda_init	Initial value for lambda (precision of the weights)
n_iter	Maximum number of iterations
normalize	When set to True, it transforms the numeric features by scaling them to a given range
tol	Stop the algorithm if w has converged.
verbose	Verbose mode when fitting the model.

When the model has been created by using default parameters, we can enhance the model with tuned parameter with the objectvice we have highest R2 and lowest RMSE value. To enhance the model, we can use function "tune_model" on PyCaret. This function tunes the hyperparameters of a given estimator. The output of this function is a score grid with CV scores by fold of the best selected model based on optimize parameter. Metrics evaluated during CV can be accessed using the get_metrics function. Custom metrics can be added or removed using add_metric and remove_metric function.[14]



```
In [13]: tunned_br_reg = tune_model(br_reg)
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	30.1314	1931.5189	43.9490	0.9976	0.0203	0.0142
1	29.3742	2956.0953	54.3700	0.9963	0.0199	0.0136
2	33.3977	4173.1007	64.5995	0.9958	0.0207	0.0134
3	33.5816	3393.4510	58.2533	0.9953	0.0264	0.0157
4	24.6004	1255.1332	35.4279	0.9983	0.0187	0.0126
5	33.4914	2828.3110	53.1819	0.9959	0.0229	0.0156
6	30.2540	2463.0093	49.6287	0.9968	0.0221	0.0149
7	24.9123	1416.5722	37.6374	0.9980	0.0172	0.0121
8	26.9381	1563.8296	39.5453	0.9977	0.0183	0.0134
9	32.3348	3068.8386	55.3971	0.9968	0.0257	0.0152
Mean	29.9016	2504.9860	49.1990	0.9968	0.0212	0.0141
Std	3.2681	903.3894	9.1893	0.0010	0.0029	0.0012

Figure 7. Parameters Model Tuning

Compared to the model that uses the default parameters, there are changes in the parameter values used, namely the values of the parameter's alpha_1, alpha_2, fit_intercept, lambda_1 and normalize. The details can be seen in the image below.

[14] M. Ali, "PyCaret," 2020, https://pycaret.readthedocs.io/en/stable/api/regression.html?highlight=tune_model#pycaret.regression.tune_model.

```
BayesianRidge(alpha_1=1e-06, alpha_2=1e-06, alpha_init=None,
              compute_score=False, copy_X=True, fit_intercept=True,
              lambda_1=1e-06, lambda_2=1e-06, lambda_init=None, n_iter=300,
              normalize=False, tol=0.001, verbose=False)
```

Figure 8. Parameters before Tunned

```
BayesianRidge(alpha_1=0.2, alpha_2=0.0005, alpha_init=None, compute_score=False,
              copy_X=True, fit_intercept=False, lambda_1=0.01, lambda_2=1e-06,
              lambda_init=None, n_iter=300, normalize=True, tol=0.001,
              verbose=False)
```

Figure 9. Parameters after Tunned

Moreover, we can see the model have lowest RMSE value after tunned parameters, comparing the default parameters, decrease by 0.03.

Evaluation

After tuned the models, we trying to test the models with new data which have been prepared before (named with unseen_data). For testing, we will used function which called predict_model. Predict_model functing require at least two parameters with one parameter for the model that have been created and the others for the data to be predicted. The using of the predict_model presented below.

```
In [15]: #prediksi dengan data test
         unseen_prediction = predict_model(tunned_br_reg, data=data_unseen)
         unseen_prediction
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Bayesian Ridge	28.2954	2061.1801	45.4002	0.9972	0.0184	0.0129

Out[15]:

	Terakhir	Pembukaan	Tertinggi	Terendah	Label
0	1325	1330	1350	1320	1323.423732
1	1325	1350	1365	1315	1331.845814
2	1295	1295	1350	1290	1313.840065
3	1275	1305	1315	1275	1285.669725
4	1250	1275	1300	1220	1258.080309
...	...	...	...	...	...
379	1400	1400	1410	1385	1384.265783
380	1165	1160	1175	1140	1149.029198
381	2420	2590	2630	2410	2526.615671
382	1390	1400	1415	1390	1389.202847
383	1220	1235	1240	1220	1218.002870

Figure 10. Testing Model with New Data

Based on figure 10, we used the tuned Bayesian rigde regression to predict the new dataset and the model have the R2 score equal to 0.9972 (near to 1 is good model) and RMSE

value is 45.4, where this can be interpreted the Bayesian rigde regression can be used for stock price prediction.

CONCLUSION

This study concludes that the Bayesian ridge regression contained in the PyCaret library can be used to predict stock prices with an R2 value of 0.9968 when modeling and 0.9972 when testing with new data and has an RMSE value of 49.199 when modeling and 45.400 when testing with new data. In addition, utilizing the PyCaret library can make it easier to create machine learning models because it uses short code and has facilities for evaluating the models created. suggestions for further model development are adding datasets and optimizing other metrics to produce better metric values.

REFERENCE

Abraham, A., and A. K. Tyagi. *Data Science for Genomics*. Elsevier Science, 2022.

Ali, M. "PyCaret," 2020. https://pycaret.readthedocs.io/en/stable/api/regression.html?highlight=tune_model#pycaret.regression.tune_model.

———. "PyCaret: An Open Source, Low-Code Machine Learning Library in Python." PYCARET, 2020. https://www.pycaret.org.

Bell, J. *Machine Learning - Hands-On for Developers and Technical Professionals*. New Jersey: Wiley, 2020.

Cannataro, M., P. H. Guzzi, G. Agapito, C. Zucco, and M. Milano. *Artificial Intelligence in Bioinformatics - From Omics Analysis to Deep Learning and Network Mining*. Elsevier Science, 2022.

Fakhruddin, H. M. *Go Public*. Jakarta: Elex Media Komputindo, 2013.

George, N. *Practical Data Science with Python - Learn Tools and Techniques from Hands-on Examples to Extract Insights from Data*. Birmingham: Packt Publishing, 2021.

Handini, S. *Buku Ajar : Manajemen Keuangan*. Surabaya: Scopindo Media Pustaka, 2020.

Haw, S. C., and K. S. Muthu. *Proceedings of the International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022)*. Atlantis Press, 2022.

Radecic, D. *Machine Learning Automation with TPOT - Build, Validate, and Deploy Fully Automated Machine Learning Models with Python*. Birmingham: Packt Publishing, 2021.

Setianto, B. *Saham-Saham Industri Consumer Goods Di BEI per Laporan Keuangan Q4 2018*. BSK Capital, 2019.

Spillane, James J. *Ekonomi Farmasi*. Jakarta: Grasindo, n.d.

Tambunan, A. P. *Menilai Harga Wajar Saham*. Jakarta: Elex Media Komputindo, 2008.

Xiao, P. *Artificial Intelligence Programming with Python*. New Jersey: Wiley, 2022.