Jurnal Mantik, 5 (4) (2022) 2720-2728
Published by:Institute of Computer Science (IOCS)
Jurnal Mantik
Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Comparison Analysis of SVM Algorithm with Linear Regression in Predicting used Car Prices

Yennimar[1], Kelvin[2], Suwandi[3], Amir[4]

[1,2,3,4]Computer Science, Universitas Prima Indonesia, Medan
Jl. Sikambing, Medan, Indonesia

Email : yennimargucci@gmail.com, kelvinwj98@gmail.com, suwandixu09@gmail.com, amirchang54@gmail.com

ARTICLEINFO

Article history:

Received: Jan 03, 2022
Revised: Jan 17, 2022
Accepted: Feb 20, 2022

Keywords:

Prediction,
Price,
Used Car,
SVM Method,
Linear Regression Method

ABSTRACT

During the high activity , car has become a basic need. On the other hand, the price of new car is getting higher. To meet these needs, people are looking for alternatives by buying used cars. One of the factors to consider when looking for a used car is price. In this study, two algorithms that are quite popular in terms of prediction will be tested, namely the Support Vector Machine algorithm and the Linear Regression algorithm in predicting used car prices. Support Vector Machine is a supervised learning method that analyzes data and recognizes patterns for regression. Support Vector Machine has the ability to solve linear and nonlinear problems. Linear Regression Algorithm is a modeling and analysis of numerical data consisting of one or more independent variables and the value of the dependent variable, with the aim of using regression analysis to estimate the value of the dependent variable based on the value of the independent variable. The result of this research is that the SVM method can perform better than linear regression. SVM can perform kernel-tricks that can handle non-linear data, thus making the non-linear data appear to be linear. but this cannot be done by Linear regression.

1. Introduction

The data mining concept hugely applied in many field such as medical, marketing analysis, education and manufacturing. Data mining can be used to process classification, prediction, estimation and gaining information from large amounts of data. Data mining is an information discovery by searching for information patterns that contain trends in very large amounts of data and helping the data storage process in a decision in the future. In data mining, the function is one part of the minor function, one of which is prediction. Prediction is a way to look for possible outcomes in the future.

Along with the high activity, car has become a basic need. On the other hand, the price of new cars is getting higher with various features embedded in new products. To meet these needs, people are looking for alternatives by buying used cars. There are many things that must be considered when buying a used car, this is because the car has been used previously by someone else and the value of the quality of the goods is certainly experiencing depreciation, one of the factors that must also be considered is the issue of price. In this case, the researcher will test two algorithms that are quite popular in terms of prediction, namely, the Support Vector Machine algorithm and the Linear Regression algorithm in predicting used car prices.

Support Vector Machine is a method that belongs to supervised learning that analyzes data and recognizes patterns for classification and regression. Support Vector Machine has the ability to solve linear and nonlinear problems The kernel approach developed in SVM can also be used to overcome the number of diverse and many classes, SVM for prediction is called SVM Regression consisting of linear and nonlinear functions. For the next algorithm, namely the Linear Regression algorithm, this algorithm is a modeling and analysis of numerical data consisting of one or more independent variables and the value of the dependent variable. The relationship between the dependent and independent variables depends on several equations, as follows: linear, exponential and the last multiple relationships, with the aim of using regression analysis to estimate the value of the dependent variable based on the value of the independent variable.

Several studies that have been tested out previously include "Analysis and Comparison of Data Mining Algorithms in Predicting GGRM Stock Prices", where the results obtained from comparisons of the Neural Network, Linear Regression, Support Vector Machine, Gaussian Process, and Polynomial Regression algorithms, it is concluded that GGRM stock price data can be predicted using the Neural Network algorithm model with the smallest prediction accuracy of RMSE 612,474 +/- 89,402 (micro: 618,916 +/- 0.000) compared to other algorithm models, so this prediction can help predict GGRM stock prices in the capital market. Meanwhile, when calculating the processing time, the K-Nearest Neighbor method is superior with a processing time of 0.0160s without K-Fold Cross Validation and 0.1505s with K-Fold Cross Validation. Other studies include "Comparison of the performance of the Neural Network Algorithm, Linear Regression, and Random Forest in the simulation of predicting the mortality rate of COVID-19 patients in Indonesia", where the results of the study show that the random forest algorithm is the algorithm that is closest to the original value in making predictions with RMS values. 94.991 +/- 0.000 followed by the neural network algorithm with an RMS value of 285.956 +/- 0.000 the furthest deviation occurred in the linear regression algorithm, which was 3285 positive deviating from the actual value.

2. Method

The steps of the work procedure in this study can be described as follows:

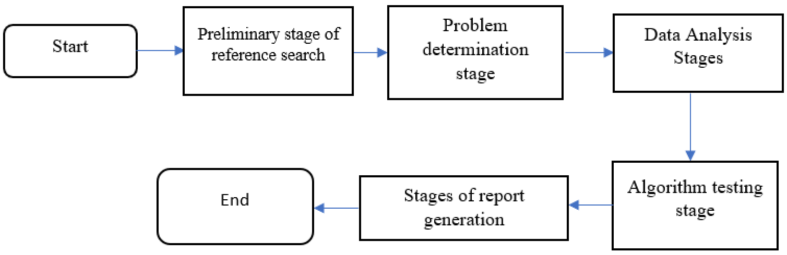

Figure 1. Research Stage

a. Introduction Stage
   The research begins with searching for references related to the research to be carried out, the references collected are in the form of a collection of previous studies.
b. Problem Determination Stage
   Determine the formulation of the problem that occurs in a used car price maker. As well as determining the problem boundaries of this research which aims to focus the scope of the research.
c. Data Collection Stage
   This stage is done by collecting data from the internet regarding the list of cars and their prices.
d. Analysis Stage
   This process analyzes datasets related to used car sales, the data will go through a selection phase regarding which features affect sales, and which features are unrelated.
e. Testing Stage
   At this stage, the data that has been collected will be tested into the SVM and Linear Regression algorithms.
f. Report Stage
   At this stage, the data that has been collected will be tested into the SVM and Linear Regression algorithms.

SVM (Support Vector Machine) is a learning system that uses a hypothetical space in the form of linear functions in a high-dimensional feature space, trained with a learning method based on optimization theory by implementing a learning bias derived from statistical learning theory. The theory underlying SVM (Support Vector Machine) itself has been developing since the 1960s, but was only introduced by Vapnik, Boser and Guyon in 1992 and since then SVM (Support Vector Machine) has grown rapidly. SVM (Support

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Vector Machine) is a relatively new technique compared to other techniques, but it has better performance in various fields of application.

Regression analysis is a statistical technique for modeling and investigating the relationship of two or more variables, the most commonly used and the simplest is simple linear regression. In Regression analysis there are one or more independent variables that can be represented by notation x and one response variable that can be represented by notation y. As the name implies, the relationship between these two variables is linear.

3. Result and Discussion

3.1 Result

The work process of the comparative analysis between the SVM and Linear Regression methods in carrying out the car price prediction process can be seen in the following diagram:

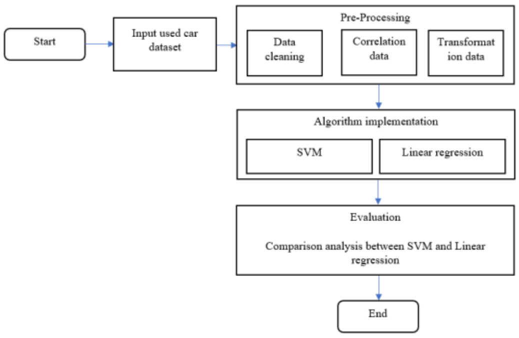

Figure 2. Work Diagram of Comparative Analysis of SVM and Linear Regression Methods

a. Input Dataset

As shown in Figure 2, the working process of the comparative analysis of the SVM and Linear Regression methods will start by inputting the used car dataset into the Jupiter Notebook application. An example of the display of dataset input results can be seen in the following image:

	Unnamed: 0	Name	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price
0	0	Maruti Wagon R LXI CNG	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	35936250.0
1	1	Hyundai Creta 1.6 CRDi SX Option	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	256687500.0
2	3	Maruti Ertiga VDI	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	123210000.0
3	6	Nissan Micra Diesel XV	2013	86999	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0	71872500.0
4	12	Maruti Swift VDI BSIV	2015	64424	Diesel	Manual	First	25.2 kmpl	1248 CC	74 bhp	5.0	114996000.0

Figure 3. Example of the Dataset Input Results to the System

b. Pre-Processing

After that, the process will proceed to the pre-processing stage which is the initial stage that must be done in the data mining method.

1). Data Cleansing

The first step of the pre-processing process is the data cleaning process, namely the process of removing data elements that are not needed in the data mining process. In Figure 3 it can be seen that there is an 'Unnamed' column that is not needed, so the 'Unnamed' element is removed from the dataset. The description of the dataset used is as follows:

The data type of each element in the dataset:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

TABLE 1.
DATA TYPE OF EACH ELEMENT

Element	Data-type
Year	Int64
Kilometers_Driven	Int64
Fuel_Type	Object
Transmission	Object
Owner_Type	Object
Mileage	Object
Engine	Object
Power	Object
Seats	Float64
Price	Float64

Dataset description :

	Year	Kilometers_Driven	Seats	Price
count	2059.000000	2059.000000	2059.000000	2.059000e+03
mean	2013.455075	59285.758621	5.300631	1.511112e+08
std	2.520335	37942.867971	0.743913	1.405421e+08
min	2001.000000	1048.000000	5.000000	1.232100e+07
25%	2012.000000	37272.500000	5.000000	6.981900e+07
50%	2014.000000	55598.000000	5.000000	1.049338e+08
75%	2015.000000	72001.500000	5.000000	1.652041e+08
max	2019.000000	720000.000000	8.000000	1.078088e+09

Figure 4. Dataset Description

After that, it is checked whether there is null data or not. The results obtained are that there is a null value in the Power column, then the data record will be deleted that has a Power element value = null.

2). Check Data Correlation (Data Correlation)

The next stage will be followed by the process of checking data correlation. Correlation is one of the analytical methods in statistics that can be used to find the relationship between two variables with quantitative properties. Correlation statistics is a method or way to find out whether there is a linear relationship between variables. If a relationship is found later, then changes that occur in one variable (X) will cause changes in the other variable (Y). At this stage, we will check the correlation between various features and prices. The stronger the correlation between features and prices, the better the dataset.

The correlation result obtained can be seen in the following figure:

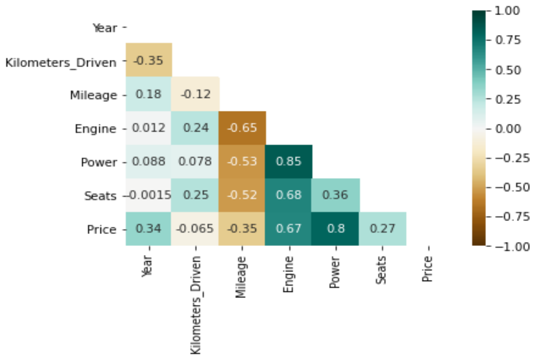

Figure 5. Correlation Result

The interpretation of the correlation coefficient obtained is as follows [21] :

TABLE 2.
DATA TYPE OF EACH ELEMENT

Correlation	Interpretation
0.00-0.199	Very Low
0.20-0.399	Low
0.40-0.599	Middle
0.60-0.799	Strong
0.80-1.000	Very Strong

Based on table 2, from Figure 7 can be obtained information that:
a). Price has a low correlation with year, milleage and seats.
b). Price has a very low kilometres_driven correlation.
c). Price has a strong correlation with the engine.
d). Price has a very strong correlation with power.

3). Data Transformation
Finally, a data transformation process will be carried out, namely the process of converting non-numeric data into numeric data to simplify the calculation process of the method used. The stages contained in the data transformation process can be seen in the following figure:

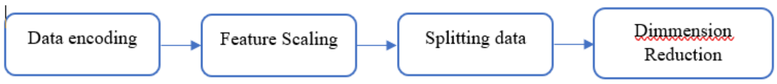

Figure 6. Transformation Data Stage

The details of data transformation used are as follows:
a). Data encoding, converting categorical data to numeric.

- Owner_Type
  Owner_Type will use label encoding technique, as shown in the following table:

TABLE 3.
LABEL ENCODING

Element	Value
First	0
Second	1
Third	2
Forth&Fifth	3

- Name, Fuel_type dan Transmission
- Name, Fuel_Type and Transmission will use a one-hot encoding technique, as shown in the following details:
  The name consists of Audi A4 2.0 TDI, Audi A4 2.0 TDI 177 Bhp Premium Plus, ..., Volkswagen Vento Petrol Highline. The process of transforming Name into Name_Audi A4 2.0 TDI, Name_Audi A4 2.0 TDI 177 Bhp Premium Plus, …, Name_ Volkswagen Vento Petrol Highline.
  Fuel Type = CNG, Diesel and Petrol. The process of transforming the Fuel Type into Fuel_Type_CNG, Fuel_Types_Diesel and Fuel_Type_Petrol.
  Transmission = Automatic and Manual. The process of transforming Transmission into Automatic Transmission and Transmission_Manual.

b). Feature Scaling After that, the process of scaling (feature scaling) will be carried out on the dataset used. This is necessary so that all features have the same range of values so that no feature has a more dominant value than the other features.
For example, the year feature contains the values 2010, 2011, 2012, … while the kilometers_driven feature contains the values 41000, 64424, 72000, 86999, … while the power feature contains the data 58.16, 63.1, 74, …. Each of these features has a different range value and because the kilometers_driven feature has a higher value than the year and power features, the year and power feature values will have no effect on the prediction

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

results obtained. To prevent this, it is necessary to perform a feature scaling process. The process will be carried out using the following command:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
scaled_df = pd.DataFrame(scaler.fit_transform(encoded_df.astype(float)))
scaled_df.columns = encoded_df.columns
```

Figure 7. Scaling

c). Splitting Dataset x and y , Before performing dimension reduction, it is necessary to separate the target data (y) from the input data (x), which can be done using the following command:

```
scaled_x = scaled_df.drop(columns=['Price'])

scaled_y = scaled_df['Price']
```

Figure 8. Scaling-2

d). PCA (Principal Component Analysis) Dimension Reduction, PCA is a dimension reduction technique. Dimensionality reduction is the process of reducing the dimensions of large-dimensional data into small-dimensional data. There are two techniques in this dimensionality reduction, namely feature selection and feature extraction. Feature selection will select influential features from the original data set, while feature extraction will form new features based on old features with fewer dimensions than before.
PCA is included in the feature extraction. This process will be carried out using the following command:

```
from sklearn.decomposition import PCA
pca = PCA(n_components = 152-1).fit(scaled_x)
pc_train = pca.transform(scaled_x)
def explained_variance(s,n_top_components):
exp_variance = np.square(s[:n_top_components]).sum()/np.square(s).sum()
def create_transformed_df(train_pca,scaled_df,n_top_components):
df_transformed = pd.DataFrame(train_pca ,columns=scaled_x.columns.values)
pca_scaled_df = create_transformed_df(pc_train, scaled_df, 88)
```

Figure 9. Scaling-3

After this stage, there will be 2043 pieces of data that will be processed further. The final result of the dataset can be seen in the following image:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	Name_Audi A4 2.0 TDI	Name_Audi A4 2.0 TDI 177 Bhp Premium Plus	Name_Audi A4 2.0 TDI Multitronic	...	Name_Mahindra XUV500 W10 2WD
0	-0.245895	-0.354632	-0.414409	0.025684	0.448911	-0.116319	1.364626	0.204152	0.566224	-0.165244	...	0.000924
1	0.532566	-0.397721	-0.094772	-0.079000	-0.207864	-0.134804	-0.003943	-0.023356	-0.022774	-0.064049	...	-0.447312
2	0.563150	-0.503300	0.194714	-0.162617	0.525630	0.246504	0.007749	-0.137355	-0.175026	0.243072	...	0.001134
3	0.459623	-0.506459	-0.269651	-0.032234	-0.026123	0.085651	0.025528	0.025180	-0.051996	-0.023616	...	0.003609
4	0.453509	-0.539982	-0.397292	-0.050033	0.054097	-0.048163	-0.005461	-0.014672	-0.083217	0.025258	...	-0.000923
...	...	...	...	...	...	...	...	...	...	...	...	...
2038	-0.862086	-0.035669	-0.154867	-0.137297	0.109191	-0.072336	-0.022193	-0.034817	-0.011947	0.044147	...	-0.002480
2039	0.696518	-0.319751	0.701707	-0.203660	0.008081	0.260857	0.080712	-0.083895	-0.043156	-0.077468	...	0.001970
2040	0.495130	-0.498240	-0.325276	-0.049675	-0.022156	-0.122956	0.013485	-0.026363	-0.023108	0.075701	...	-0.265005
2041	0.458246	-0.597669	-0.530565	-0.018901	0.128941	0.193417	-0.365187	-0.189378	0.804709	0.066881	...	0.002217
2042	-0.871015	0.007706	-0.022438	-0.116823	0.032585	0.060683	-0.020995	0.013486	-0.037681	-0.034615	...	0.000206

Figure 10. Dataset after Pre-processing-1

Name_Mahindra XUV500 W6 2WD	Name_Mahindra XUV500 W8 2WD	Name_Mahindra XUV500 W8 4WD	Name_Maruti A-Star Vxi	Name_Maruti Alto 800 2016-2019 LXI	Name_Maruti Alto 800 LXI	Name_Maruti Alto K10 2010-2014 VXI	Name_Maruti Alto K10 VXI	Name_Maruti Alto LXi
1.579210e-16	-2.476910e-16	-1.689297e-17	-1.155862e-16	9.266759e-18	-2.524655e-16	2.056051e-16	8.223815e-16	-7.559029e-17
-2.272178e-02	-3.328291e-01	-1.596653e-01	3.065088e-01	-3.636698e-02	-1.660151e-01	-2.505025e-02	-7.942396e-03	-2.902679e-01
-2.926717e-16	-1.976577e-16	7.822615e-16	-4.838981e-16	5.268129e-16	-9.953145e-16	-4.020975e-16	3.073963e-16	1.099131e-16
4.581621e-17	-5.077396e-16	-2.050130e-16	7.294813e-17	9.833621e-17	-6.980262e-16	-3.229507e-16	3.692709e-15	2.599667e-16
-2.848654e-16	3.956570e-17	1.731568e-16	-4.947401e-16	1.013720e-16	1.561082e-16	8.102298e-17	-1.294621e-15	-1.659079e-16
...	...	...	...	...	...	...	...	...
1.663612e-15	-1.710750e-15	-1.784501e-15	-4.907712e-16	2.130558e-15	3.541448e-15	2.686631e-15	-2.095810e-16	-4.487921e-16
3.707601e-15	4.246743e-16	5.897072e-16	3.791268e-16	1.867754e-15	1.335070e-15	2.638656e-15	1.200345e-15	-5.805068e-16
-1.311138e-01	-5.631439e-02	4.031642e-02	1.424881e-01	4.525677e-01	5.588271e-02	-1.060312e-01	2.465669e-01	3.661240e-01
6.338936e-16	-8.543094e-17	1.693604e-16	-1.582158e-16	-2.356204e-16	6.369279e-17	9.094526e-17	1.679010e-15	1.499346e-16
-4.101328e-18	-8.584452e-17	2.482681e-16	-1.455732e-16	-4.758861e-16	-4.458550e-16	-3.921217e-16	-7.482492e-17	2.057341e-17

Figure 11. Dataset before Pre-Processing-2

c. SVM Method

After the pre-processing stage, the process will continue with the training and testing stages. The dataset obtained will be divided into two parts, namely training data and testing data, with a division scale of 7: 3, which means that the amount of training data is 70% and the amount of testing data is 30%.

Support vector machine is a model that can be used for regression and classification. In classification modeling, SVM has a more mature and clearer mathematical concept than other classification techniques. SVM can also solve classification and regression problems with linear and non linear. SVR is an SVM method which is applied to the regression case. SVR aims to find a function f(x) as a hyperplane (separation line) in the form of a regression function which corresponds to all input data by making the smallest possible error.The parameters used in this SVM method include the value of C= 100, the value of gamma = 0.1, the value of epsilon = 0.05 and the type of kernel used is RBF (Radial Basis Function), as shown in the following formula:

$$K\ (x, x') = \exp(-\gamma \|x - x'\|2) \qquad (1)$$

Description:

K = kernel

γ = gamma

x, x’ = input x and y

d. Linear Regression Method

After the pre-processing stage, the process will continue with the training and testing stages. Regression or prediction models involving more than one independent variable or predictor. The term multiple regression can also be referred to as multiple regression. The word multiple means plural or more than one variable. The working process of the Linear Regression method can be done using the following command:

```
from sklearn.linear_model import LinearRegression
lr_reg = LinearRegression().fit(train_x,train_y)
lr_score = lr_reg.score(train_x,train_y)
print('Linear-regression Data-train accuracy: {}'.format(lr_score))
test_lr_score = lr_reg.score(test_x,test_y)
```

Figure 12. Training and Test Linear Regression-1

The results obtained from the Linear Regression method can be detailed, as follows:

```
lr_score = lr_reg.score(train_x,train_y)
print('Linear-regression Data-train accuracy: {}'.format(lr_score))
```

Figure 13. Output Training

Output from training: Linear-regression Data-train accuracy: 0.8484492144508385

```
test_lr_score = lr_reg.score(test_x,test_y)
print('Linear regression test-score accuracy: {}'.format(test_lr_score))
```

Figure 14. Output Testing

Output from testing: Linear regression test-score accuracy: 0.8363942356919476

3.2 Discussion

The SVM method can perform training with better results than linear regression. Basically, both SVM and linear regression are doing the tuning on the hyperplane line and fitting the distribution of the dataset to get the minimum loss.

The difference is that the SVM method can overcome the spread of non-linear data, for example output_x : 1 ,2 ,3 ,4 ,5 and output_y: 1 ,4 ,7, 5, 6.

From the data above, it can be seen that the value of y increases with the increase of x, but at the value of x = 4 , the value of y = 5 where y decreases so that in this dataset there is a non-linear part. SVM can do a kernel-trick that can handle the non-linear data, thus making the non-linear data appear to be linear. but it can't be done by Linear regression.

4. Conclusion

From the discussion in the previous chapters, finally the research in this final project can be drawn several conclusions, among others. The SVM method can perform training with better results than linear regression. The accuracy of the test results with the SVM method is better than the Linear Regression method. The SVM method can perform a kernel-trick that can overcome the non-linear distribution of data.

5. References

[1] F. Ginting, E. Buulolo, and E. R. Siagian, "Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 3, no. 1, pp. 274–279, 2019, doi: 10.30865/komik.v3i1.1602.

[2] A. A. Suryanto, "Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," Saintekbu, vol. 11, no. 1, pp. 78–83, 2019, doi: 10.32764/saintekbu.v11i1.298.

[3] M. Electric, "Prediksi Harga Mobil Bekas Dengan Machine Learning," vol. 6, no. 5, 2021, [Online]. Available: https://emea.mitsubishielectric.com/ar/products-solutions/factory-automation/index.html.

[4] F. Destiawati, H. Dhika, N. Network, and I. Pendahuluan, "Model Penentuan Pembelian Kondisi Mobil Bekas," 2017.

[5] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil," J-SAKTI (Jurnal Sains Komput. dan Inform., vol. 2, no. 2, p. 152, 2018, doi: 10.30645/j-sakti.v2i2.78.

[6] F. R. Lumbanraja, R. S. Sani, D. Kurniawan, and A. R. Irawati, "Implementasi Metode Support Vector Machine Dalam Prediksi Persebaran Demam Berdarah Di Kota Bandar Lampung," J. Komputasi, vol. 7, no. 2, 2019, doi: 10.23960/komputasi.v7i2.2426.

[7] A. Bode, "Perbandingan Metode Prediksi Support Vector Machine Dan Linear Regression Menggunakan Backward Elimination Pada Produksi Minyak Kelapa," Simtek J. Sist. Inf. dan Tek. Komput., vol. 4, no. 2, pp. 104–107, 2019, doi: 10.51876/simtek.v4i2.57. [8] F. H. Hamdanah and D. Fitrianah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikra, Kecil, dan Menengah," J. Nas. Pendidik. Tek. Inform., vol. 10, no. 1, p. 23, 2021, doi: 10.23887/janapati.v10i1.31035.

[9] R. Maulana and D. Kumalasari, "Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham Ggrm," J. Inform. Kaputama, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: https://finance.yahoo.com/quote/GGRM.J.

[10] M. F. Fibrianda and A. Bhawiyuga, "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM)," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. II, no. 9, pp. 3112–3123, 2018.

[11] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," J. Inform., vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.

[12] M. R. Redo and A. Irianti, "Perbandingan performa Algoritma Neural Network, Regresi Linier, dan Random Forest dalam simulasi prediksi angka kematian pasien COVID-19 di Indonesia," Semin. Nas. Has. Penelit. dan Pengabdi. Masy., pp. 54–62, 2021, [Online]. Available: https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2915.

[13] A. Darmawan, N. Kustian, and W. Rahayu, "Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuya," STRING (Satuan Tulisan Ris. dan Inov. Teknol., vol. 2, no. 3, p. 299, 2018, doi: 10.30998/string.v2i3.2439.

[14] Suhardjono, W. Ganda, and H. Abdul, "Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso," Bianglala Inform., vol. 7, no. 2, pp. 97–101, 2019.

[15] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi," JUITA J. Inform., vol. 7, no. 2, p. 71, 2019, doi: 10.30595/juita.v7i2.4378.

[16] P. D. Atika and Rasim, "Implementasi Jaringan Syaraf Tiruan Metode Backpropagation untuk Prediksi Penjualan Mobil Bekas," Implementasi Jar. Syaraf Tiruan Metod. Backpropagation untuk Prediksi Penjualan Mob. Bek., vol. 18, no. 2, pp. 107–112, 2019, doi: 10.36054/jict-ikmi.v18i2.70.

[17] H. P. Tambunan and S. Zetli, "Implementasi Metode K-Nearest NeigborDalam Peramalan Penjualan Mobil Bekas di Kota Batam," Comasie, vol. 3, no. 3, pp. 21–30, 2020.

[18] A. Dirgantara, S. Herdyansyah, and R. Rasenda, "Klasifikasi Penerimaan Mobil Bekas Berdasarkan Metode Neural Network," J. Ris. Inform., vol. 2, no. 1, pp. 43–48, 2020, doi: 10.34288/jri.v2i1.119.

[19] B. J. Putra, T. B. Kurniawan, D. Antoni, and A. H. Mirza, "Prediksi Kebutuhan Alat Kesehatan Rumah Sakit Menggunakan Metode Algoritma Regression Linier dan Naïve Bayes," vol. 11, no. 2, 2020.

[20] A. Saiful, "Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 8, no. 1, pp. 41–50, 2021, doi: 10.35957/jatisi.v8i1.701.

[21] N. Fitri, M. Mawardi, and R. A. Kurniawan, "Korelasi Antara Keterampilan Metakognisi Dengan Aktivitas Dan Hasil Belajar Siswa Pada Mata Pelajaran Kimia Kelas X Mia Sma Negeri 7 Pontianak," AR-RAZI J. Ilm., vol. 5, no. 1, 2017, doi: 10.29406/arz.v5i1.655.

[21] Siregar, S. D., Panjaitan, B., Girsang, E., & Dabukke, H. (2019). Learning Media Using Discovery Learning Approach to Improve Student Learning Outcomes. Jurnal Pendidikan Fisika, 8(2), 120-125.

[22] Dharshinni, N. P., Azmi, F., Fawwaz, I., Husein, A. M., & Siregar, S. D. (2019, July). Analysis of accuracy K-means and apriori algorithms for patient data clusters. In Journal of Physics: Conference Series (Vol. 1230, No. 1, p. 012020). IOP Publishing.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.