Journal of Software Engineering, Information and Communication Technology (SEICT)

Journal homepage: https://ejournal.upi.edu/index.php/SEICT

Prediction of Mobile Phone Ratings with SVM Regression Model

Arya Dwi Ramdhani[1], Fahri Admana Budi[2], Rizki Dimulya[3]*

Software Engineering, Universitas Pendidikan Indonesia, Indonesia
Correspondence: E-mail: aryaramdhani@upi.edu

ABSTRACT

Mobile is a communication tool with capabilities such as computers that are easy to carry anywhere with various functions for human life. Mobile phones certainly have quite interesting trends, such as the emergence of models, types, and brands, which vary. This study aimed to determine the prediction of mobile phone ratings based on various criteria using the SVM method. These criteria include price, camera, internal memory, storage, color, etc. The regression type gets predictive results from the SVM model, where values are adjusted to the model. Although the accuracy is not good, in the prediction process the difference is not too far, but slightly different. The more you add related features, the training accuracy will be better.

ARTICLE INFO

Article History:
Submitted/Received 05 Nov 2023
First Revised 14 Nov 2023
Accepted 28 Nov 2023
First Available online 01 Dec 2023
Publication Date 01 Dec 2023

Keyword:
SVM,
Rating, Mobile,
Regression.

1. INTRODUCTION

Mobile phones are communication tools with capabilities like computers that are easy to carry anywhere with various multifunctions for human life. Mobile phones have features such as games, music, camera, video, and social media, which can be used anywhere and anytime. Cellphones certainly have interesting trends, such as the emergence of various models, types and brands. The criteria for cell phones are seen from the following criteria: brand, model, color, internal memory, RAM, rating, selling price and original price. Remember that RAM and internal memory are important for a cellphone to get a high rating.

RAM and internal memory on cellphones, of course have various sizes. RAM is where commands or applications on the cellphone run, where all applications will run on it. RAM and internal memory are very influential on cellphones because RAM can help the cellphone run several applications simultaneously. In contrast, if full, internal memory is a storage area that will slow down the performance of the cellphone. Therefore, the role of RAM and internal memory is important in a cellphone.

RAM and internal memory on a cellphone sometimes also affect the price of a cellphone. The higher the RAM and the larger the internal memory, the higher the cellphone price. However, it will not always be like that; it also involves other criteria for a cellphone. These criteria include price, internal memory, camera and battery capacity. These criteria will provide the public's assessment of a cellphone, generally called a rating.

The SVM algorithm model is an algorithm from the classification method that can produce a learning process in a classification problem, which is translated as an effort to find a hyperplane line to separate the two groups (Darmawan, Kustian, & Rahayu, 2018). The SVM model used is Regression Analysis. Regression analysis is a statistical process for estimating the relationship between a dependent or criterion variable and one or more independent variables or predictors (Ghorakavi, 2021). Based on the Journal "Choosing Mobile Phones with Conjoint Analysis based on Preferences of High School Students in Bekasi Regency for Online Schools" written by Feri Prasetyo, Agus Dendi Rachmatsyah, Wahyu Tisno Atmojo, obtained from 60 respondents with a price perception of 28.112% in first place then RAM 19.931% Then the camera is 16.972%. Thus, this research aims to determine cellphone rating predictions based on various criteria using the SVM method.

2. METHODS

The method used in this research was based on the application created. This application can predict cellphone ratings based on various criteria, which uses the Support Vector Machine (SVM) Regression model. The flow of research carried out consists of several stages, namely as shown in the **Figure 1**.

DOI: https://doi.org/10.17509/seict.v4i2.64392
p- ISSN 2774-1656 e- ISSN 2774-1699

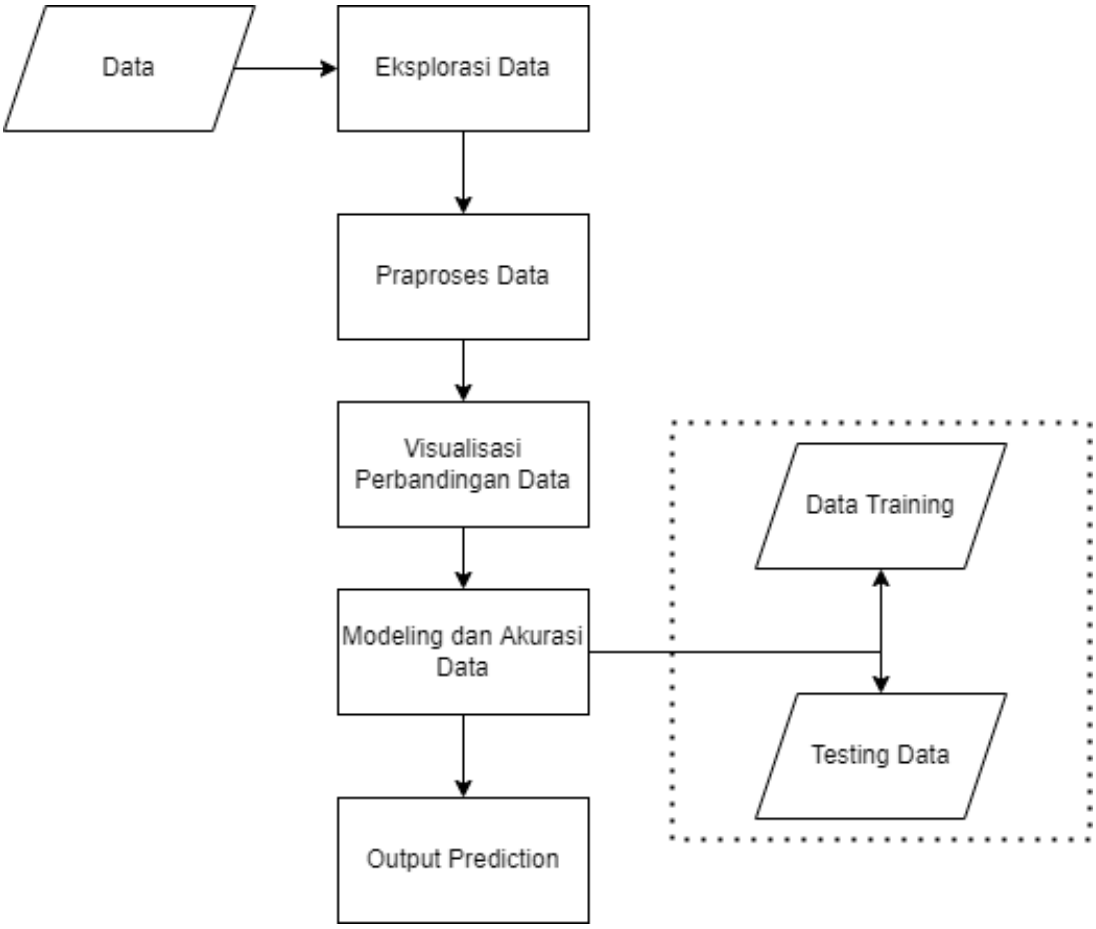

Figure 1. Research Method

Starting from obtaining data, then exploring the data, and then the data goes through the data pre-processing stage. After that, the data is visualized in comparison to obtain modeling and data accuracy. Of course, during the modeling process and obtaining data accuracy, training data and testing data are needed. So that, in the end, it produces a predicted value output.

3. RESULTS AND DISCUSSION

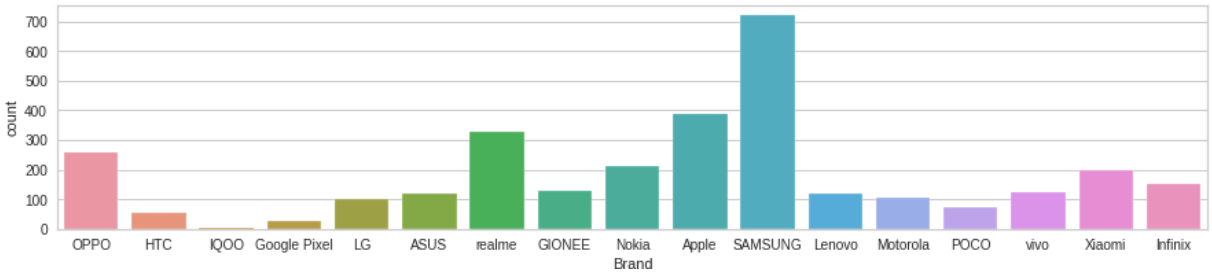

Figure 2. Number of Datasets

Exploratory Data Analysis (EDA) is needed to analyze existing data in **Figure 2**. Exploratory data analysis (EDA) is a strategy for analyzing data to find hidden information that can provide new insights for a basis for decision-making (Wahyuni, E. D., Arifiyanti, A. A. and Kustyani, M.,

DOI: https://doi.org/10.17509/seict
p- ISSN 2774-1656 e- ISSN 2774-1699

2019). Exploratory Data Analysis is an initial investigative testing process that aims to identify patterns, find anomalies, test hypotheses and check assumptions. By carrying out EDA, users will be greatly helped in detecting errors from the start, you can identify outliers, know the relationship between data, and explore important factors from the data (Khasanah, 2020).

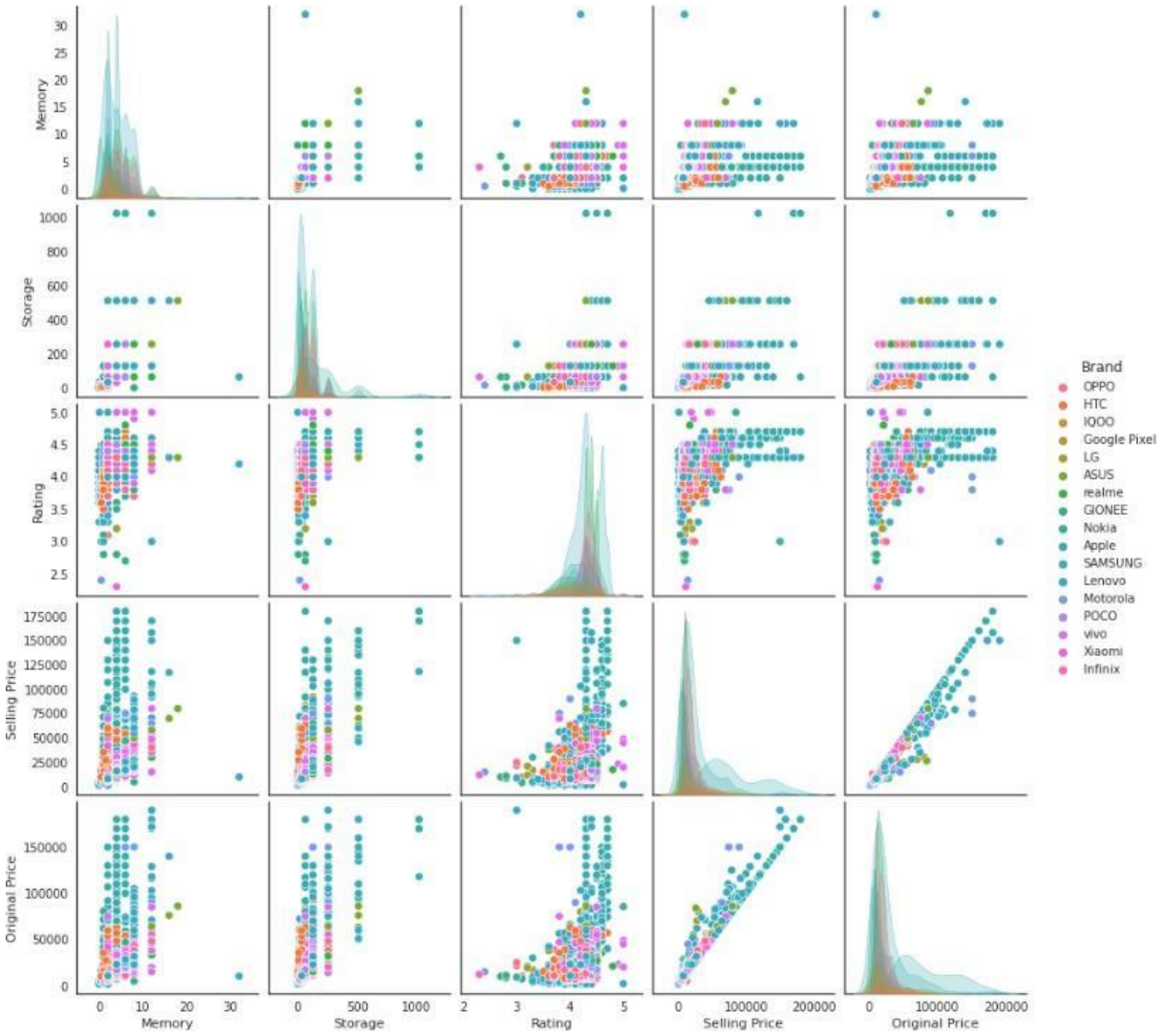

Figure 3. Correlation of Data

As the picture shows in **Figure 3**, a comparative relationship exists between one feature and another. These features are the criteria that influence cell phone rating predictions.

DOI: https://doi.org/10.17509/seict.v4i2.64392
p- ISSN 2774-1656 e- ISSN 2774-1699

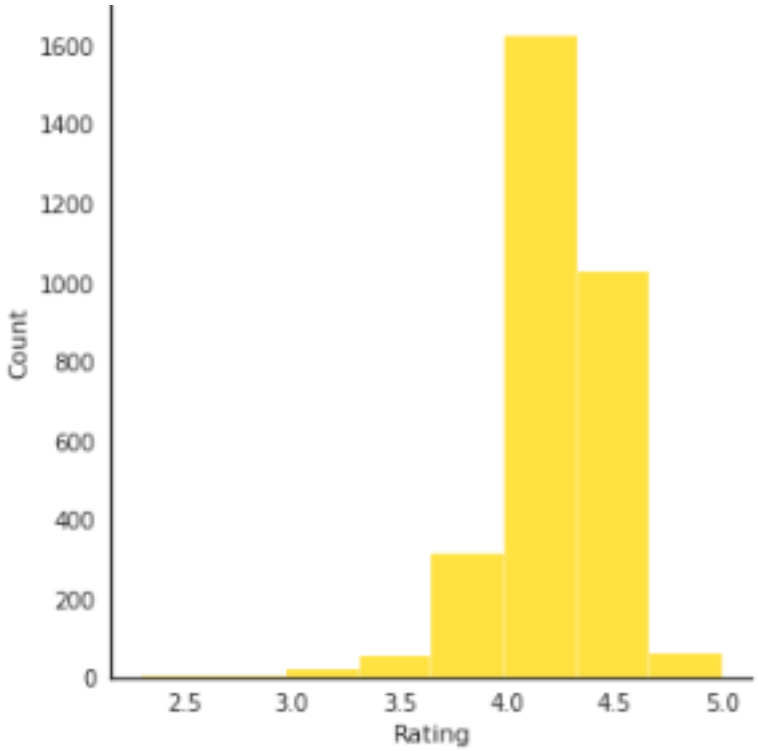

Figure 4. Rating value and the total number

Figure 4 shows the relationship between the rating value and the total number, with the highest rating in the range of 4.0 - 4.5. The boxplot displayed in **Figure 5** shows that some cell phones, in terms of ratings, still have many outliers.

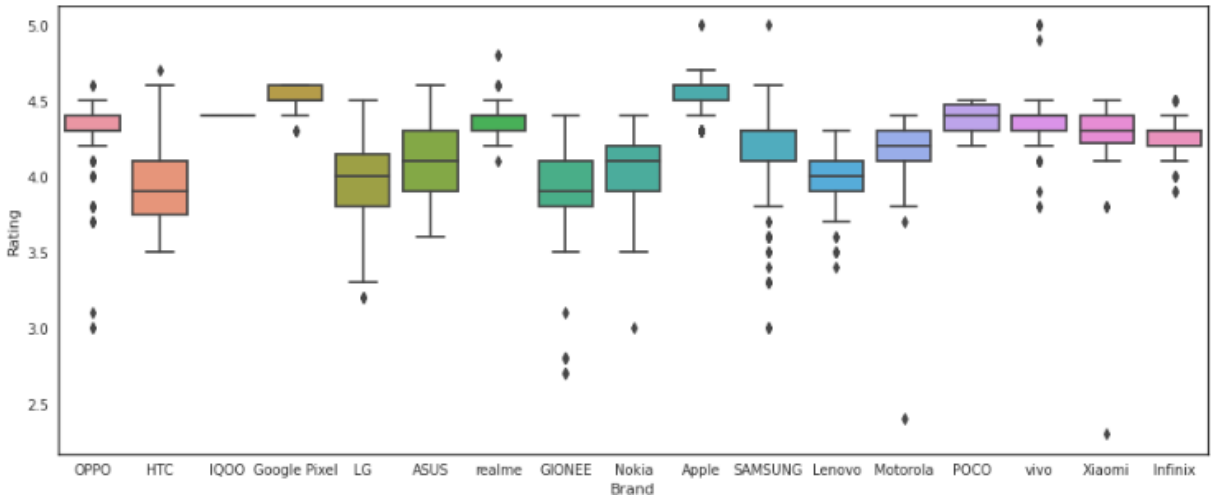

Figure 5. Outliers of Ratings

Based on the EDA that was obtained, the research was carried out using the SVM method with initial results as shown in **Figure 6**.

DOI: https://doi.org/10.17509/seict
p- ISSN 2774-1656 e- ISSN 2774-1699

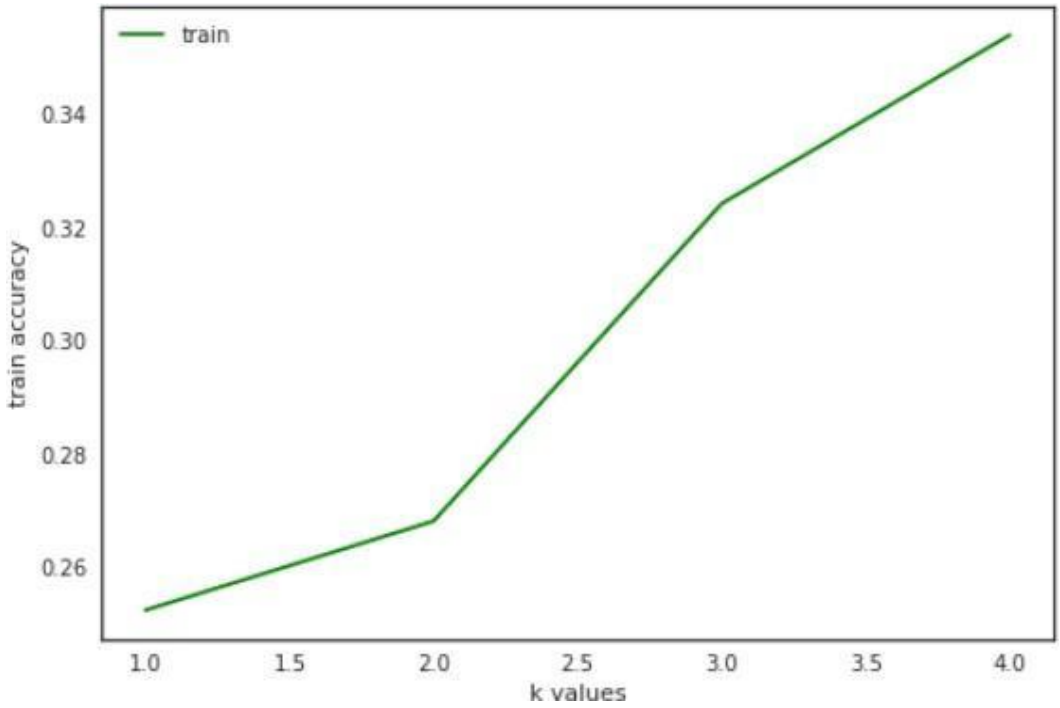

Figure 6. Initial Result SVM Method.

Figure 6 shows the k value selected with a training accuracy of 35% and a test accuracy of 43%. This means that the test accuracy value is higher than the training accuracy. This data also shows that the more criteria, such as memory, storage, etc., are included, the higher the training accuracy will be. Therefore, other features were added to increase accuracy.

After adding new features to the calculation, the k value and training accuracy are obtained as in **Figure 7**.

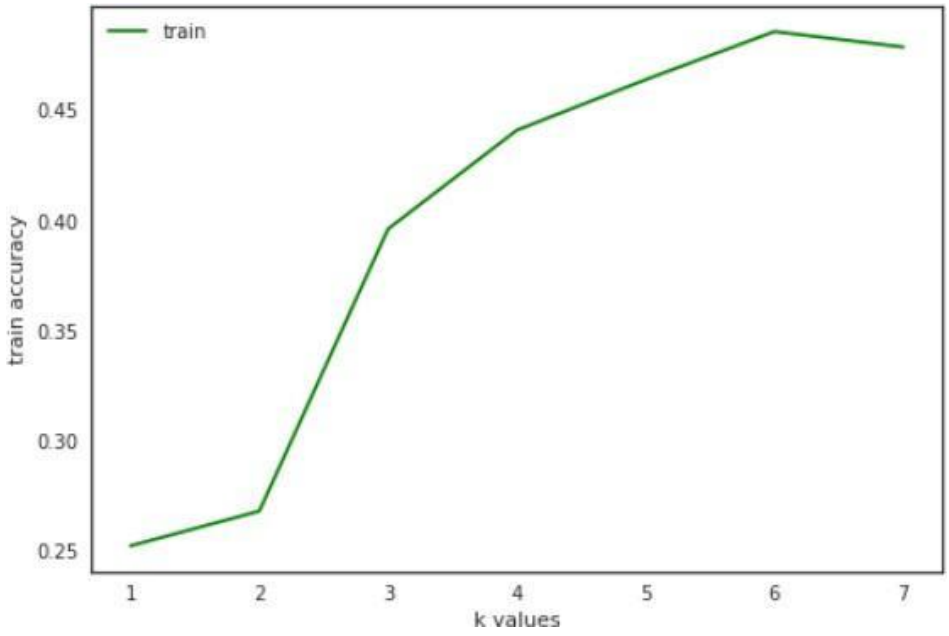

Figure 7. Training Accuracy

DOI: https://doi.org/10.17509/seict.v4i2.64392
p- ISSN 2774-1656 e- ISSN 2774-1699

Figure 7 shows the k value selected with a training accuracy of 48% and a test accuracy of 53%. This means that the test accuracy value is higher than the training accuracy. Because the accuracy has increased, SVM fitting was carried out on the new data, with a training accuracy of 49% and a test accuracy of 52%. However, after adding more criteria to 7 criteria, the training accuracy reached its maximum point at the 6th criterion, while at the 7th criterion it decreased. Therefore, the 7th criterion, namely the color criterion, has been removed so that the criterion remains 6.

Because the test accuracy value decreases, we immediately proceed to SVM prediction with the SVM test value, namely the training target data. So, we get the top 10 SVM test results. The prediction results obtained are as follows

	Memory	Storage	Selling Price	Original Price	Brand Label	Model Label	Color Label	Actual Rating	Predicted Rating
235	0.124945	0.062498	0.100559	0.095186	0.6875	0.254107	0.766458	4.3	4.249117
364	0.124945	0.031248	0.122968	0.116397	0.4375	0.292442	0.931034	4.0	4.176835
326	0.062441	0.031248	0.050302	0.047614	0.4375	0.844469	0.053292	3.9	4.027252
717	0.124945	0.062498	0.050302	0.076715	0.9375	0.805038	0.120690	4.4	4.354927
2204	0.062441	0.031248	0.055892	0.052905	0.5000	0.834611	0.409091	4.1	4.043977
1065	0.062441	0.015623	0.032091	0.030376	0.6250	0.058050	0.808777	4.0	4.083576
2287	0.015563	0.003904	0.017887	0.016931	0.5000	0.128149	0.072100	3.6	3.936277
1657	0.249953	0.124998	0.273890	0.343912	0.8125	0.465498	0.197492	4.0	4.382449
2101	0.124945	0.031248	0.125769	0.235451	0.8125	0.481928	0.815047	4.4	4.243208
2929	0.124945	0.062498	0.050302	0.063487	0.3750	0.511501	0.731975	4.3	4.199730

Figure 8. Prediction Result

The table shows the rating prediction results. For example, in the Actual Rating, the rating value is 4.3, while the Predicted Rating shows a value of 4.249117. This means that the predicted value is not completely accurate because the predicted value is almost close to the actual value but has not yet reached the same value.

4. CONCLUSION

From the model created, conclusions can be drawn from the regression type SVM model to get prediction results, where values are adjusted into the model. The actual value is 4.3, but the predicted value is 4.2, which means that this prediction is only close to the actual value and does not exactly match the actual value. Even though the accuracy is not good, in the prediction process the difference is not too big, but slightly different. Because the furthest difference is only 0.3, apart from that, the more related features you add, the better the training accuracy will be.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

DOI: https://doi.org/10.17509/seict
p- ISSN 2774-1656 e- ISSN 2774-1699

6. REFERENCES

Darmawan, A., Kustian, N., and Rahayu, W. (2018). IMPLEMENTASI DATA MINING MENGGUNAKAN MODEL SVM UNTUK PREDIKSI KEPUASAN PENGUNJUNG TAMAN TABEBUYA. Jurnal String Vol. 2 No. 3, 299-307.

E. D. Wahyuni, A. A. Arifiyanti, and M. Kustyani. (2019). "Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining," Pros. Nas. Rekayasa Teknol. Ind. dan Inf. XIV Tahun 2019, vol. 2019, no. November, pp. 263–269.

Feri Prasetyo, RACHMATSYAH, A. D., and Wahyu Tisno Atmojo. (2021). Pemilihan Handphone dengan Analisis Konjoin Berdasarkan Preferensi Siswa SMA di Kabupaten Bekasi untuk Sekolah Online. Teknomatika, 11(02), 137-142.

Ghorakavi, Veena. (2021). ML | Types of Regression Techniques. Url : https://www.geeksforgeeks.org/types-of-regression-techniques/

Khasanah, Latifah Uswatun. (2020). Exploratory Data Analysis : Pahami Lebih Dalam untuk Siap Hadapi Industri Data. Url : https://www.dqlab.id/data-analisis-machine-learning-untuk-proses-pengolahan-data

DOI: https://doi.org/10.17509/seict.v4i2.64392
p- ISSN 2774-1656 e- ISSN 2774-1699