

Dynamic Pricing Analytic of Airbnb Amsterdam Using K-Means Clustering

Fitrianingsih

Information Systems Management,
Postgraduate Program
Gunadarma University
Depok, Indonesia
fitrianingsih@staff.gunadarma.ac.id

Dewi Agushinta Rahayu

Information Systems Management,
Postgraduate Program
Gunadarma University
Depok, Indonesia
dewiar@staff.gunadarma.ac.id

Figa Rizfa Zazila

Information Systems Management,
Postgraduate Program
Gunadarma University
Depok, Indonesia
figarizfa@gmail.com

Abstract—The development of the world of providing accommodation services in the world is increasingly showing a very rapid development. One of these developments is in the field of information technology. One of the developments in tourism technology is the Airbnb company. Airbnb is an online-based booking or rental service for private rooms, apartments or houses. Owners can rent back accommodations on a daily basis such as hotel rooms. This research will conduct a dynamic pricing analysis that can help Airbnb hosts choose the optimal property price based on the selected season. This research uses clustering method with K-Means algorithm. The data that will be processed is Airbnb data in Amsterdam, which comes from the Airbnb dataset website for one year from 04 March 2021 to 06 March 2022. This research is expected to provide an overview of one marketing strategy by giving companies the freedom to adjust the price of service products when the situation arises, suitable and possible. Dynamic pricing is one of the marketing strategies for companies to compete with competitors. The accuracy of the research using the K-Means Clustering model is 96.21%. The net income result is 24201, the actual income is 16068, the expected profit or loss is 8133. Where this amsterdam airbnb is right to use the Dynamic Pricing method, because the income generated is optimal.

Keywords—Airbnb, Dynamic Pricing, K-Means Clustering, Revenue

I. INTRODUCTION

The development of the world of providing accommodation services in the world is increasingly showing a very rapid development. One of these developments is in the field of information technology. Technology will never be consumed by the times, technology will continue to develop from time to time and experience changes from imperfect to perfect. This development will certainly be more useful for many people. One of the developments in tourism technology is the Airbnb company. For Airbnb owners it can be an income, while for tenants it can be an alternative to get experiences and accommodations that are cheaper than staying at a hotel [1]. The Airbnb concept itself is known as the sharing economy concept or also known as collaborative consumption [2].

Several studies related to dynamic pricing have been carried out by previous researchers, including research [3], which uses a Support Vector Machine (SVM) based on segmentation models and discrete choice-models to find the accurate value of predicting passenger choice and increasing airline industry revenue in Europe in April – October 2014. The advantage of this study is that the best accuracy for the model is 42% and revenue has increased by more than 20%, but the data used causes a bias in price estimates.

Research [4] uses the Kernel Principal Component Analysis (KPCA) method and K-Means Clustering analysis. The results obtained were that using the K-Means Clustering analysis which produced a sum-of-squares of 84.31 and PCA (Principal Component Analysis) of 78.17% resulted in a home energy management system providing energy and being able to observe their household energy consumption in real-time. time, but the data used is not balanced between RTP and TOU.

Implementation in determining dynamic pricing at Airbnb hotels in Tokyo and Tahoe in research [5] has been carried out. The method used is the Binary Classification Model by using the Gradient Boosting Machine to predict the price of possible orders per night, the Regression Model by using the Support Vector Regression and Customize Loss Function to train the data and the last is to apply the output logic of the two models to make the final price prediction. The results show that Airbnb price predictions since February 08, 2018 on average vary widely for Tokyo and Tahoe over the next 120 nights. Judging from these results, the model can capture dynamic markets in a timely manner.

This research will conduct a dynamic pricing analysis that can help Airbnb hosts choose the optimal property price based on the selected features and season. This research uses clustering method with K-Means algorithm. Choosing k-means as a research method based on dynamic pricing research that has been studied previously, K-means produces the greatest predictive value for related research. The data to be processed is Airbnb data located in Amsterdam, North Holland, The Netherlands which comes from the Airbnb dataset website for one year [6]. This research is expected to provide an overview of one marketing strategy by giving the company the freedom to adjust the price of service products when the situation is suitable and possible. Dynamic pricing is one of the marketing strategies for companies to compete with competitors.

II. THEORY

A. Dynamic Pricing

The definition of dynamic pricing is a strategy of applying different prices to the same product or service by adjusting the time, event, place and consumer characteristics. The implementation of dynamic pricing strategy is based on the reality that consumers are heterogeneous. Therefore, consumers usually have the maximum ability of the price of the product or service that they are willing to pay or what is called the reservation price. With dynamic pricing, it is expected to be able to discriminate prices at the individual level based on the track record of previous customers [7]

Dynamic pricing is obtained based on various conditions and situations, namely demand or demand, availability or availability, competitors and forecasts. The value of dynamic pricing will adjust to these factors, if demand, availability and prices increase, dynamic pricing will increase prices [8].

B. K-Means Clustering

Clustering or classification is a method used to divides the data series into several groups based on predetermined similarities. Cluster is a group or a collection of data objects that are similar to each other in the same cluster and are similar to objects in different clusters. Objects will be grouped into one or more clusters so that objects in one cluster will have a high similarity between one another [9].

K-Means is a non-hierarchical data clustering method that tries to partition existing data into one or more clusters or groups so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into clusters.

III. RESEARCH METHOD

This research consists of several stages of the concept-making research process in conducting dynamic pricing analysis of Airbnb prices using the unsupervised learning K-Means Clustering model. The general chart of this research method can be seen in Figure 1.

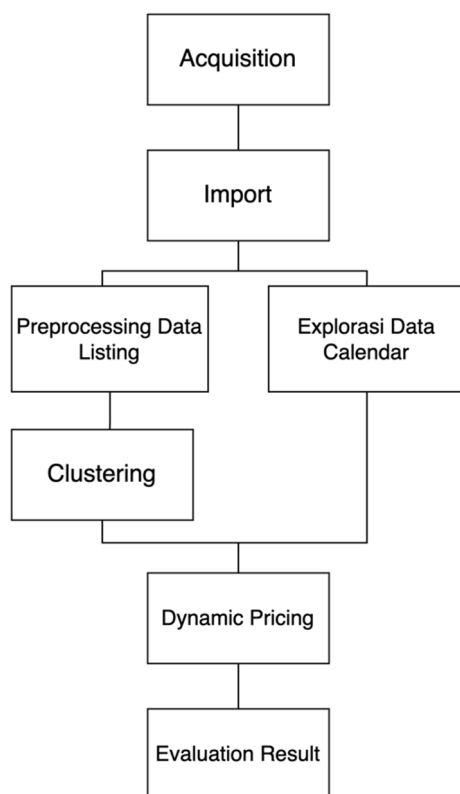


Fig. 1. Stages of Research Method

Based on Figure 1, the initial stage of the research method is the acquisition stage, namely the stage of collecting data from the Inside Airbnb site. The dataset used is Airbnb data located in Amsterdam, North Holland, The Netherlands.

The second stage of this research is the stage of importing data or entering data that has been downloaded through the acquisition stage. The data taken are listing.csv data and calendar.csv data. The Airbnb dataset is then processed using Google Collaboratory.

The third stage after importing data is the pre-processing stage. The purpose of pre-processing is to adjust the data to be more compatible with the library to be used, the data used for the pre-processing stage are listing.csv and calendar.csv data.

The fourth stage after preprocessing is clustering or grouping, with the clustering process using a K-Means algorithm, data will focus on one class and will get optimal results.

The fifth stage is data exploration by calendar.csv data. At this stage the calendar data is processed to remove irrelevant or useless data, with the aim of this stage being to see the months or seasons that can affect Airbnb accommodation prices.

The next stage is the creation of a dynamic pricing model. This stage is the core stage of programming. This stage combines two datasets that have been processed previously in the pre-processing stage, namely listing.csv and calendar.csv data exploration. Modeling is done by using test cases to test the data results.

The last stage is the process of evaluating the results, which is to see the results of making a dynamic pricing model. The results shown are the average occupancy, net income, actual income and losses on Airbnb Amsterdam for the 2021-2022 period. The explanation of each stage of this research method will be explained in the results and discussion.

IV. RESULTS AND DISCUSSION

A. Dataset

Airbnb data is available at <http://insideairbnb.com/>. Inside Airbnb Sites which are sites that contain publicly or publicly available information about Airbnb sites. Data on Inside Airbnb has been analyzed, cleaned and aggregated so as to facilitate public discussion.

The data used is Airbnb Amsterdam listing data for the period 04 March 2021 to 06 March 2022, as many as 18,201 rows and 74 columns for listing data and 6,643,081 rows and 7 columns for calendar data. Listing data is a collection of attribute lists from the name of the unit to the point of Airbnb accommodation which is used to create a pattern from the clustering stage in the study, while calendar data is a collection of attributes from date to price of Airbnb accommodation which is used to visualize unit prices and period dates and create dynamic pricing patterns.

The next step is to enter the dataset that has been downloaded from the airbnb site, which is then uploaded to Google Drive, which is then imported into the Google Collaboratory by entering the library import drive, and after that the listing.csv and calendar.csv datasets are entered. Airbnb Amsterdam. Figure 2 is a sample of the listing.csv dataset and Figure 3 is a sample of the calendar.csv dataset.

Calendar.csv data that has been entered is then seen how many days and unique listings or the number of unique elements. There are 368 days and 18198 unique listings in the calendar.csv data. The data calendar needs to be edited in the format so that the next process will be better. The features that changed the format are price and date, because these two features are very useful for determining dynamic pricing on Airbnb Amsterdam. The price feature is formatted by removing the comma (,) and dollar (\$) symbols, while the date feature is formatted from scalar to datetime object. The edited format is as shown in Figure 6.

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	451719	2021-03-05	t	190.0	\$190.00	3.0	60.0
1	230710	2021-03-05	f	125.0	\$125.00	14.0	60.0
2	230710	2021-03-06	f	125.0	\$125.00	14.0	60.0
3	230710	2021-03-07	f	125.0	\$125.00	14.0	60.0
4	230710	2021-03-08	f	125.0	\$125.00	14.0	60.0

Fig. 6. Edit Result of Listing Data Format

Next, the date and price features are combined to create a visualization plot of the monthly Airbnb Amsterdam price. It can be seen in Figure 7 that the highest initial price was in February 2022 at around \$176.5 and the lowest was in March at \$157.1.

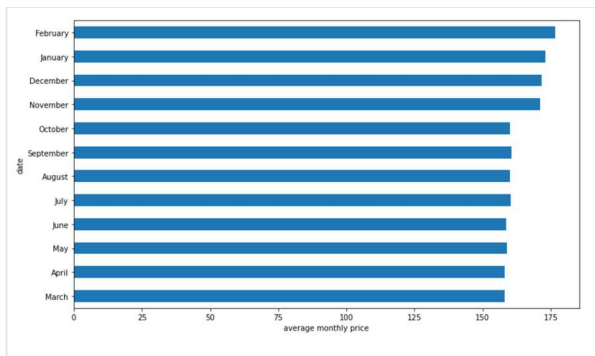


Fig. 7. Visualization Bar of Data Calendar Horizontal

Visualization of data in weekly or weekly format, starting with importing the datetime library. The library is used to create a new column called dayofweek which contains the day of the week and is denoted by the numbers 0-6. 0 for Sunday to 6 for Saturday.

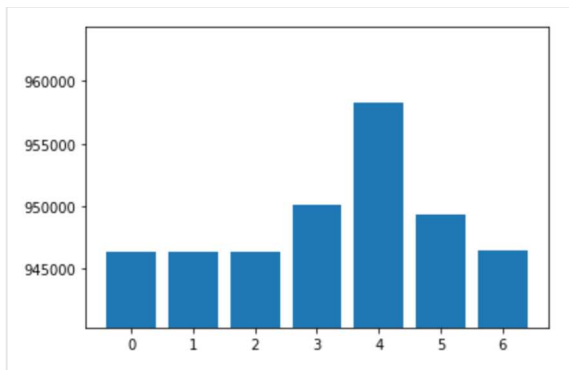


Fig. 8. Weekly Accommodation Price Bar Graph

C. Clustering

The clustering process using the K-Means algorithm is used to find cluster values for the dynamic pricing process at

a later stage. The K-means algorithm requires a K value. The K value is obtained using the Elbow method. The data used for the clustering process is listing data that has been processed at the preprocessing stage. The clustering stages are shown in Figure 9. In entering the clustering stage, the first thing to do is find the K value.

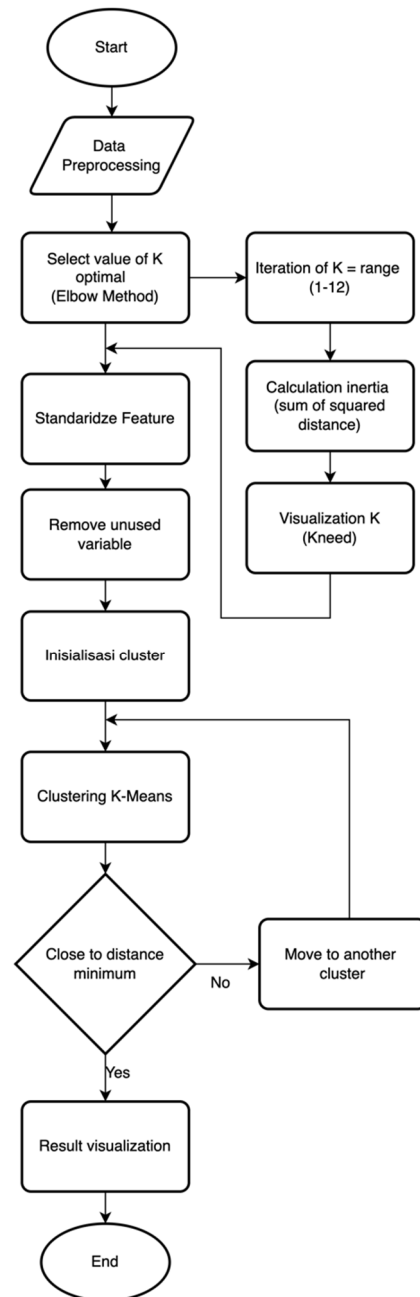


Fig. 9. Stages of Process Clustering

The first stage is carried out to determine the optimal K value, namely using the Elbow method. Furthermore, in the elbow method, the value of the K iteration is determined with a range between 1-12, where each K value is taken using an inertia sum of square distance, which is the squared distance between the mean points at each point of the cluster. After that plot or visualize the result of the value of K using the kneed library, where the visualization is viewed based on the centroid distance by identifying the knee/elbow point based on the elbow method of the line corresponding to the data. The results show that the K value obtained is (Figure 10).

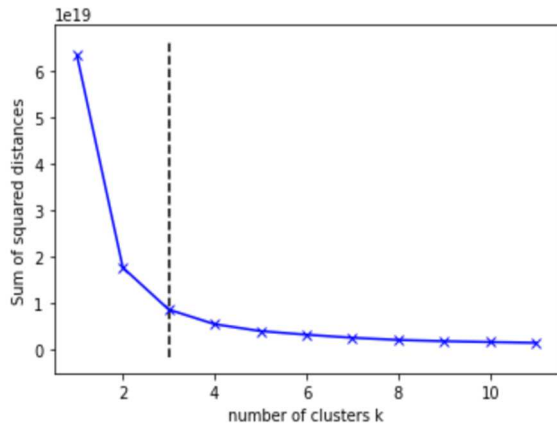


Fig. 10. Graph of K Values by Kneed

After entering the K-means value, the next step is to standardize the features with the sklearn library and the RobustScaler method with the aim of standardizing the dataset so that it is ready for the next process. The method in this standardize feature is `scaler.fit_transform(numeric)` which is used to adjust the data and convert the data to numeric. Results like Figure 11.

	id	host_id	host_total_listings_count	host_has_profile_pic	host_identity_verified	latitude	longitude	accommodates	bedrooms	beds	price
0	-0.985393	-0.394183	1.0	0.0	-1.0	0.469756	0.053056	0.5	0.0	0.0	-0.125392
1	-0.985085	-0.392921	0.0	0.0	0.0	-0.664048	-0.759854	1.0	2.0	2.0	1.053292
2	-0.984538	-0.392186	0.0	0.0	0.0	0.780833	0.048877	0.5	0.0	0.0	0.313480
3	-0.984430	-0.392029	-1.0	0.0	0.0	0.595444	0.190747	1.0	1.0	3.0	0.952978
4	-0.984383	-0.394038	0.0	0.0	-1.0	-0.342498	-0.461253	0.0	0.0	1.0	-0.877743

Fig. 11. Standardize Feature Results

The K-means algorithm will be run using a K value of 3. This value is obtained based on the process in the previous stage using the elbow method. If the k-means value is not appropriate or the clustering is not correct or not close to the minimum value, then iteration will be carried out so that the value is successfully obtained.

In this study, 2 iterations were carried out. Stage 2 iteration is carried out in the same way, namely running the K-means algorithm, previously created 3 cluster lists, each named `clus1`, `clus2` and `clus3`. The results show the best clustering using `clus1`, which is depicted in Figure 13.

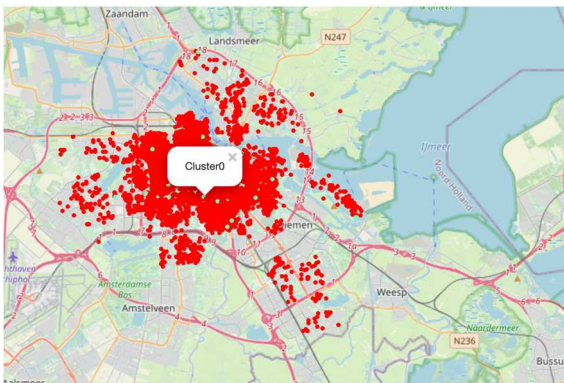


Fig. 12. Folium Map K-Means 1

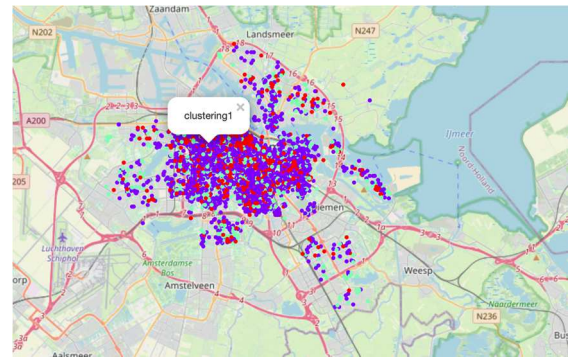


Fig. 13. Folium Map K-Means 2

D. Dynamic Pricing

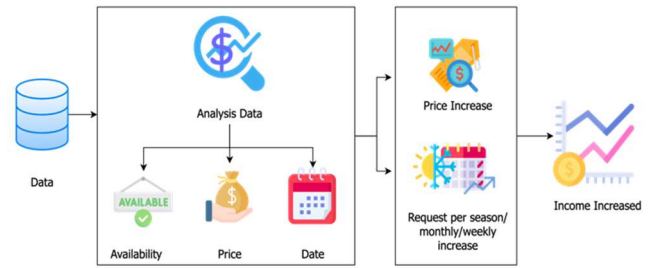


Fig. 14. Dynamic Pricing Illustrations

The data that has been inputted will be analyzed using existing data such as availability, price and date. After analyzing and processing, the demand for seasons, monthly and weekly increases, so the price will increase so that the income for Airbnb Amsterdam increases. The dynamic pricing process is depicted in Figure 15.

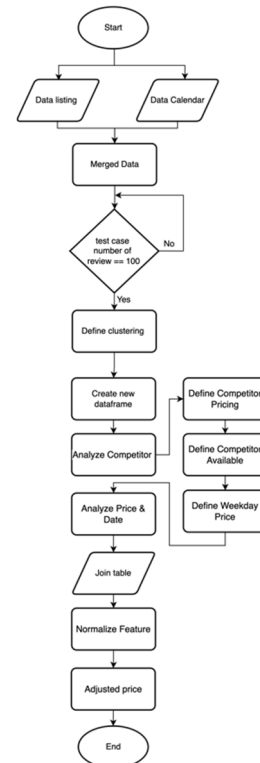


Fig. 15. Stages of the Dynamic Pricing Process

Calendar.csv and listing.csv data that have been processed in the previous process, then the two data are merged. The calendar.csv data that has the listing_id feature is renamed to id so that it can join the listing.csv data. Next, create a case sample from the merged dataset.

In this study, the selected case is the number of the number of review feature which is assumed to be one of the complete data. The selected number of review data is 100.

The feature sought in this study is that it must meet the basic characteristics of dynamic pricing, which is a value that can be updated every day and has a direct effect on prices, therefore there are two features that will support the next process, namely Availability and Price. By looking at the average results of competitor prices or competitor pricing and the average availability of competitors or competitor availability every day or per day.

```
date
2021-03-04      1
2021-03-05    530
2021-03-06    950
2021-03-07   1301
2021-03-08   1501
...
2022-03-02   1244
2022-03-03   1247
2022-03-04   1247
2022-03-05    272
2022-03-06     0
Name: available, Length:
```

Fig. 16. Mean Available Perday

```
dayofweek
0    159.911586
1    159.971178
2    160.014271
3    160.167207
4    163.058553
5    163.108619
6    159.965034
Name: price_x, dtype:
```

Fig. 17. Mean Price Day Perweek

By combining it with the date feature, the average price and availability can be described. Figure 16 is the result of the average availability per day for one year. Figure 17 is the result of the weekly average, it can be seen that Friday (4) and Saturday (5) are 1.63% larger and Monday (0) are lower prices than other days.

```
date
March      159.599806
January    160.493131
June       160.632638
February   160.647795
November    160.785348
April      160.854437
October    161.026626
May        161.052883
December   161.061330
September  161.325074
August     161.386193
July       161.789749
Name: price_x, dtype: float64
```

Fig. 18. Price by Month

The data that has been processed is then modified to get optimal results. The step taken in modifying the data is to create a weekday multiplier that is useful for weekend prices

to be higher than weekdays prices. Figure 19 is an index of data that has been carried out by dynamic pricing.

	date	available	competition_price	case_price	dayofweek	trial_price	weekday_mult	adj_price
27	2021-04-01	1836	159.723381	200.0	3.0	254.0	0.000	254.000
28	2021-04-02	1855	163.003582	200.0	4.0	249.0	0.014	252.486
29	2021-04-03	1854	163.134056	200.0	5.0	248.0	0.015	251.720
30	2021-04-04	1856	160.019970	200.0	6.0	255.0	0.000	255.000
31	2021-04-05	1837	160.023910	200.0	0.0	253.0	0.000	253.000
32	2021-04-06	1832	159.765201	200.0	1.0	253.0	0.000	253.000
33	2021-04-07	1825	159.916271	200.0	2.0	252.0	0.000	252.000
34	2021-04-08	1834	159.934181	200.0	3.0	253.0	0.000	253.000
35	2021-04-09	1846	162.763589	200.0	4.0	248.0	0.014	251.472

Fig. 19. Index Dynamic Pricing Data

E. Result Evaluation

Accuracy is obtained based on the predicted values in the y_pred and y_test datasets obtained at the normalize feature stage. The y_pred and y_test data are reshaped first to reshape the data array.

While accuracy is obtained by calculating the distribution of predictive data divided by test data multiplied by 100. The results obtained are 96.21% accuracy, where accuracy is quite good because it is close to 100%.

The net revenue model or net income is obtained based on the reduction of the expected value and the non-booking expectation value, in which the two values must first be found using the functions discussed in chapter 3. The net income result is 24,201.

Actual income is obtained based on the reduction in the number of prices and the addition of non-bookings. The actual income result is 16,068.

Expected profit or loss is obtained based on the calculation of the value of Expected Net Revenue minus Net Revenue. The result of the expected profit or loss is 8133.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

A dynamic pricing process on Airbnb using the Airbnb Amsterdam dataset for the period 04 March 2021 – 06 March 2022 downloaded through the Inside Airbnb website, was carried out using the unsupervised learning model K-Means Clustering. Dynamic pricing is one of the marketing strategies for companies to compete with competitors.

Model creation was carried out using data of 18,201 rows and 74 columns for listing data and 6,643,081 rows and 7 columns for calendar data. The results of the study in the form of clusters using the K-Means Clustering model, namely accommodation enthusiasts per month with the highest price being in July in the summer of 161,789 and the lowest in March in the spring of 159,599. In addition, the highest prices of accommodation on Airbnb Amsterdam per week are on Fridays and Saturdays. Where Friday is the day before the weekend, where people choose to take a vacation and stay at a hotel or accommodation instead of at home. The results of the accuracy of the study using the K-Means Clustering model were 96.21%. The net income was 24201, the actual income was 16068, the expected profit or loss was 8133. Where airbnb amsterdam is right to use the Dynamic Pricing method, because the income generated is optimal.

B. Recommendations

Headings Based on the results of Analisis dynamic pricing on Airbnb Amsterdam conducted, researchers can provide suggestions, namely six data changes that can generate optimal income, using other methods in determining dynamic pricing, and can be applied using a website or mobile application to facilitate the input or output produced and other researchers, the results of the study can be used as a reference in developing models for prediction using the K-Means Clustering method.

REFERENCES

- [1] Apa itu Airbnb dan bagaimana cara kerjanya?. Airbnb. Available <https://www.airbnb.co.id/help/article/2503/apa-itu-airbnb-dan-bagaimana-cara-kerjanya> [Access: January 26, 2022].
- [2] Oskam, J., & Boswijk, A. (2016). Airbnb: the future of networked hospitality businesses. *Journal Of Tourism Futures*, 2(1), 22-42. doi: 10.1108/jtf-11-2015-0048.
- [3] Delahaye, T., Acuna-Agost, R., Bondoux, N., Nguyen, A.Q., & Boudia, M. (2017). Data-driven models for itinerary preferences of air travelers and application for dynamic pricing optimization. *Journal of Revenue and Pricing Management*, 16, 621-639.
- [4] Koolen, D., Sadat-Razavi, N., & Ketter, W. (2017). Machine Learning for Identifying Demand Patterns of Home Energy Management Systems with Dynamic Electricity Pricing. *Applied Sciences*, 7, 1160.
- [5] Ye, P., Qian, J., Chen, J., Wu, C., Zshou, Y., Mars, & Zhang, L. (2018). Customized Regression Model for Airbnb Dynamic Pricing. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- [6] Inside Airbnb. Adding data to the debate. Inside Airbnb. Retrieved 27 January 2022, Available: <http://insideairbnb.com/get-the-data.html> [Access: January 27, 2022].
- [7] Li, W., Hardesty, D.M., & Craig, A.W. 2018. The impact of dynamic bundling on price fairness perceptions. *Journal of Retailing and Consumer Services*, 40 204–212.
- [8] Popovic, A. (2022). Dynamic Pricing: Benefits, Strategies and Examples. Retrieved 12 September 2022, from <https://www.price2spy.com/blog/dynamic-pricing-explained/>
- [9] Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1-5. doi: 10.46576/djtechno.v1i1.964.
- [10] Ban, G., & Keskin, N.B. (2020). *Personalized Dynamic Pricing with Machine Learning: High Dimensional Features and Heterogeneous Elasticity. Econometrics: Econometric & Statistical Methods - Special Topics eJournal*.
- [11] Jadhav, R., Vaisha, A., Nawarangee, M., & Dhruv, A. (2020). Dynamic Pricing Intelligence Approach in E-Commerce Market using KNN. *International Journal Of Management And Applied Science (IJMAS)*, 6(5), 25-28.
- [12] Shukla, A., Bhattacharya, P., Tanwar, S., Kumar, N., & Guizani, M. (2020). DwaRa: A Deep Learning-Based Dynamic Toll Pricing Scheme for Intelligent Transportation Systems. *IEEE Transactions on Vehicular Technology*, 69, 12510-12520.
- [13] Hosen, M., & Amin, R. (2021). Significant of Gradient Boosting Algorithm in Data Management System. *Engineering International*, 9(2), 85-100. <https://doi.org/10.18034/ei.v9i2.559>
- [14] Gao, F., Li, C., & Zhang, B. (2021). A Dynamic Clustering Algorithm for Lidar Obstacle Detection of Autonomous Driving System. *IEEE Sensors Journal*, 21(22), 25922-25930. doi: 10.1109/jsen.2021.3118365
- [15] Harahap, Z., Leonandri, D., Pratiyudha, D., & Julvirta, E. (2020). *Hotel Revenue Management (Manajemen Pendapatan Hotel)* (1st ed.). Malang: Intelligensia Media.