Analysis of Tourism Businesses Number in the Entertainment and Recreation Sector using Predictive Apriori Algorithm

Romarta Yemima Manurung¹, Dewi Purwita Sari², Nabila Agustinah³, Razel Syanahieskara⁴, Imam Tahyudin [ID]⁵*, **Nurfaizah⁶, Siti Alvi Sholikhatin⁷**

1,2,3,4,5,6,7Department of Information System, Universitas Amikom Purwokerto, Purwokerto, Indonesia
*Corresponding author: imam.tahyudin@amikompurwokerto.ac.id

Abstract:

Data Analysis of the Number of Tourism Businesses in the Entertainment and Recreation Sector is used as data sources for extracting information. This research conducts data mining on the number of tourism businesses in the entertainment and recreation sector to support decision-making information. This research analyzes the tourism business number in the entertainment and recreation sectors. The method uses using predictive Apriori algorithm.Knime software is used to test data and process data and the number of tourism businesses in the entertainment and recreation sector at the domestic level by using business data whose numbers are increasing or decreasing. From entering nodes 1, 2 and 3 to getting node 4, the final result. The results show the data set that produces the final result for every tourism business data. The result showed that the tourism number in entertainment and recreation sectors is increasing. Furthermore, the prediction result of entertainment and recreation with the best accuracy are balls, discotheques, massage parlors, karaoke, live music, massage parlors, sports, physical fitness centers, and family recreation facilities and spas.

Keywords: Predictive A priori, Data Mining, Tourism Businesses, Knime, Entertainment

1. INTRODUCTION

Tourism is a sector that plays an essential role in increasing income. Indonesia has beautiful nature and diverse cultures, so the tourism sector needs to be strengthened. Tourism is a sector considered highly profitable and has the potential to develop as an asset to be used as a source of national and state income. To extract information from vast databases, the science of data mining combines learning strategies, pattern recognition, statistics, databases, and visualization. There are two types of learning in data mining: supervised learning and unsupervised learning [1]. Each teaching has its method and algorithm. Methods included in Supervised learning include regression, classification, predictive, and summarization.

Meanwhile, unsupervised learning consists of clustering, association, and knowledge discovery methods [2]. Identifying the dataset's accuracy was the goal of this research. This study compares the grouping of tourism data using two supervised and unsupervised methods with a priori predictive algorithms.

The use of a priori algorithm is used because it is an algorithm that can be parallelized on large databases [3][4][5][6].

2. THEORY

A. Data Mining

Finding knowledge in a sizable data set is done through data mining. Data mining is examining a large amount of data to discover something new, accurate, and practical before a pattern is found in the data. [7]. Knowledge discovery database (KDD) is the entire process of transforming raw data into intriguing data patterns and information consumers need, as knowledge includes data mining as a key component.

B. Data Mining Grouping

Based on the tasks that can be performed, data mining is split into several areas [8] [9]:

1. *Description*
    Many researchers want to experiment with methods or techniques for expressing the trends and patterns seen in the data.
2. *Estimation*
    The model is created using a full report with a target variable as a temporary conjecture. Almost estimation is equal to classification.
3. *Predictions*
    The difference between estimations, classifications and our projection of the value of the outcomes in the future is that our prognosis is nearly the same.
4. *Classification*
    In the classification, there is a category target variable. For example, the category of business types, sub-districts, and regions.
5. *Clustering*
    It is a collection of accounts, observations, or paying close attention to and staging related objects. A cluster is a group of reports that are similar and do not differ from other cluster reports in any way.
6. *Association*
    Finding qualities that arise within a year is the task of data mining. Alternatively, it is known as shopping cart analysis in the corporate world.

C. Predictive Apriori Algorithm

Finding associative rules between a group of things is done using the data mining technique known as association rule mining. Association analysis is one of the data mining methods that form the foundation of another method. The rules of the association are usually expressed in the form [8]:

{A, B} => {C} (support = 10%, confidence = 50%). There are many association rule mining algorithms. One of them is the Predictive Apriori algorithm.

Predictive A priori developed by Scheffer. This method uses a larger support value and is exchanged for a higher confidence value, and accuracy is calculated in the Bayesian Network. This algorithm's output maximizes the value of accuracy for the data that needs to be dated. According to the user's specifications, this algorithm generates the desired number of rules. An algorithm is defined as follows:

Suppose D is a database where r is an individual record generated by the static process P; suppose $X \rightarrow Y$ becomes an association setting. Predictive accuracy $c(X \rightarrow Y) = Pr$ (r meets Y | r meets X) is the conditional probability of $Y \subseteq r$ given that $X \subseteq r$ when the distribution r is set by P.

Mining association rules that can help forecast depressive symptoms and identify undergraduate students who need special care have been made possible by the Predictive Apriori algorithm employed in Youth Suicide Prevention by Screening Depressive Symptoms. [10]. Similar research was also conducted using an a priori predictive algorithm to strengthen the data to be carried out to predict data [11]. Similar research was also conducted to apply an a priori predictive algorithm and successfully mapped student profiles to improve college promotion strategies [12]. This study also used the Apriori method to identify association rules from historical drug sales data. The study's findings include the development of the five biggest rules with a propensity to buy particular pharmaceuticals utilizing the a priori algorithm method [13]. Additionally, analyzing data on crimes against women and traffic accidents was done using a priori predictive algorithms, with the latter producing better outcomes [14] [15].

D. *Knime Application*

Knime analytics is a cornerstone that combines a large number of nodes for input/output data processing, pretreatment and cleaning, modeling, analysis, and data mining, as well as a variety of interactive features, including scatter plots, parallel coordinates, and other tools. The software incorporates all of the well-known Weka analysis modules for data mining and other plugins that make it possible to execute R-scripts and give users access to a statistical library. KNIME is based on the Eclipse platform and uses a modular API that is easily expandable when necessary, and it can be extended to understand and give first-level support for highly domain-specific data. Custom nodes and types can be developed in KNIME within a few hours.

3. METHOD

3.1 Research Design

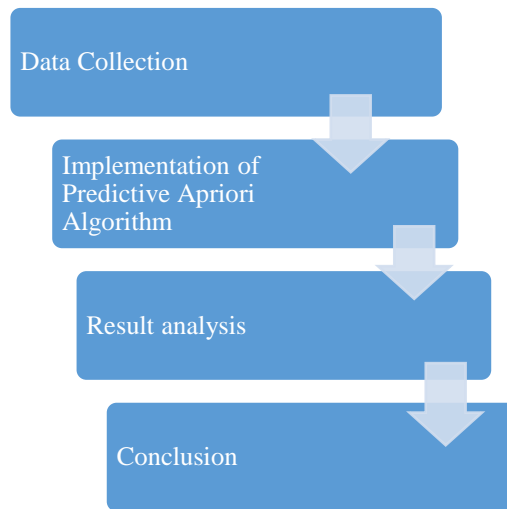The research design in this study is as follows:

Fig 1. Research Design

The research design is presented as follows:
1. The beginning of this study is the principal stage for deciding the title and the problem to be studied.
2. Problem identification, which is an explanation of the data that has been passed in entering the data
3. In terms of collecting research data utilizing libraries, and journals, as reference material in determining the parameters used in the chaining, the source of research data was obtained from the data. Jakarta.go.id. The data used in this study is data on the number of tourism businesses in the field of entertainment and recreation.
4. Data processing in this study uses data mining techniques and a priori predictive algorithm methods.
5. The design of this study is implementing tourism number data into the predictive a priori algorithm method and using knime software.

3.2 Data Collection Methods

This study used data collection techniques, namely, library studies. Such as libraries, and journals, as reference material in determining the parameters used in research. The research data sources are obtained from data. Jakarta.go.id. Data used in this study is data on the number of tourism businesses in the field of entertainment and recreation. The total data set are 1.015.

3.3 Predictive A priori Algorithm

Data is quite essential in this day and age. Data itself can be a valuable set of information if processed more deeply. Good data processing will produce good

information as well to support the best decision-making for solving a certain problem.

4.   RESULT AND DISCUSSION

4.1 Perspectives on data analysis and processing

Based on the data that has been obtained in this study, a search and analysis process will then be carried out to group tourism data so that different patterns are obtained so that they can be used as support in making decisions.

Analysis of Experimental Processes and Results

In an initial trial of the dataset, it was found that most of the analyzes were useless. Therefore, the re-experiment was carried out using a KNIME data mining tool. This can be done by classifying the types of businesses, districts, and regions into different groups and applying additional a priori algorithms to sort the data results on the number of tourism businesses in the entertainment and recreation sector, as shown in Figure 2.
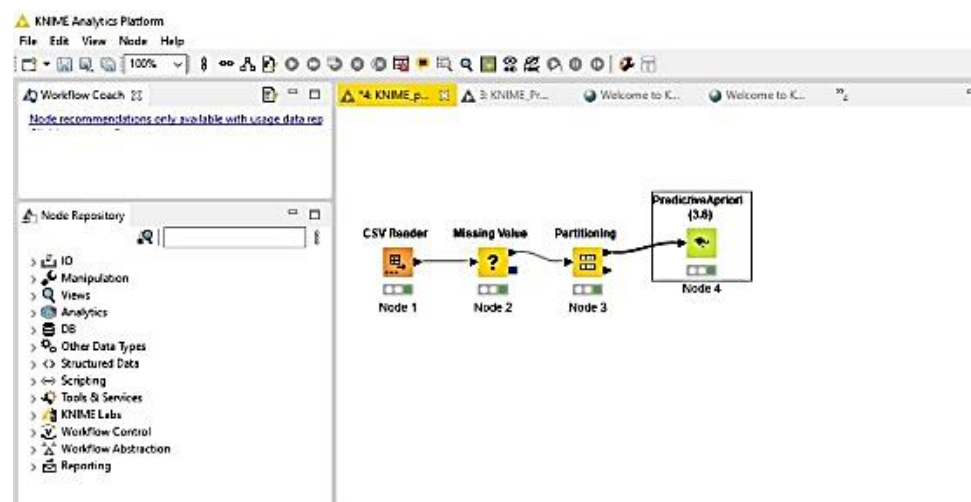

Fig 2. Implementation of KNIME

From the implementation, it is made using an a priori predictive algorithm with the minimum support = 10% and minimum confidence = 50%. The result obtained that the entertainment and recreations business are balls, discotheques, massage parlors, karaoke, live music, massage parlors, sports and physical fitness centers, family recreation facilities, and spas), districts (kebon Jeruk, Tambura, Taman Sari, Grogol Petaburan, Cempaka Putih, and Palmmerah, Kelapa Gading, Koja), regions (West Jakarta, Central Jakarta, South Jakarta, North Jakarta) with different accuracy results.

5. CONCLUSIONS

This research was conducted to apply data mining in grouping tourism and entertainment business data so that socialization can be carried out in the Jakarta area. The basic problem in this research is the scope of tourism regarding the type of business that is increasing and the number of visitors increasing. Based on the analysis, it can be concluded that the entertainment and recreation sectors over the minimum support and minimum confidence are balls, discotheques, massage parlors, karaoke, live music, massage parlors, sports, and physical fitness centers, family recreation facilities, and spa. Furthermore, they are located in districts (kebon Jeruk, Tambura, Taman Sari, Grogol Petaburan, Cempaka Putih, and Palmmerah, Kelapa Gading, Koja), regions (West Jakarta, Central Jakarta, South Jakarta, North Jakarta). Future work is better to compare the result with the other ARM algorithms.

AUTHOR CONTRIBUTIONS

Conceptualization; Romarta Yemima Manurung [R.Y.M], Imam Tahyudin [I.T], methodology; [R.Y.M],[I.T]; validation; [R.Y.M],[I.T], formal analysis; [R.Y.M],[I.T], investigation; [R.Y.M],[I.T], data curation; [R.Y.M],[I.T], writing—original draft preparation; [R.Y.M],[I.T]; writing—review and editing; [R.Y.M],[I.T]; visualization; [R.Y.M],[I.T], supervision [R.Y.M],[I.T], project administration; [R.Y.M],[I.T], funding acquisition; [R.Y.M],[I.T], have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. A. M. L. Afit Dan Y. B. Yusuf, "Analisis Sistem Pengelolaan, Pemeliharaan Dan Keamanan Jaringan Internet Pada It Telkom Purwokerto," Evolusi, Vol. 6, No. 2, 2018, Doi: 10.31294/Evolusi.V6i2.4427.

2. R. Lumbantoruan, P. Sariguna, and J. Kennedy, "Analisis Data Mining Dan Warehousing," *J. Ilm. Bul. Ekon.*, vol. 19, no. 1, pp. 1410–3842, 2015.

3. L. R. Mabrur angga, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI Program Studi Teknik Informatika Jurnal Komputer dan Informatika (KOMPUTA)," *J. Komput. dan Inform.*, vol. 1, pp. 53–57, 2012.

4. D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*. 2014. doi: 10.1002/9781118874059.

5. B. Kavšek and N. Lavrač, *APRIORI-SD: Adapting association rule learning to subgroup discovery*, vol. 20, no. 7. 2006. doi: 10.1080/08839510600779688.

6. D. Magdalene Delighta Angeline, "Association Rule Generation for Student Performance Analysis using Apriori

Algorithm," *SIJ Trans. Comput. Sci. Eng. its Appl.*, vol. 1, no. 1, pp. 12–16, 2013.

7. M. H. Santoso, "Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom," *Brill. Res. Artif. Intell.*, vol. 1, no. 2, pp. 54–66, 2021, doi: 10.47709/brilliance.v1i2.1228.

8. J. Han, M. Kamber, and J. Pei, *Data Mining Concept and Technique*, Third. USA: Elsevier, 2012.

9. Kusrini and L. Taufiq Emha, *Algoritma Data Mining Yogyakarta*, no. February. Yogyakarta: ANDI OFFSET, 2009. [Online]. Available: https://books.google.co.id/books?id=-Ojclag73O8C&printsec=frontcover#v=onepage&q&f=false

10. S. Saikin and K. Kusrini, "Model Data Mining Untuk Karekteristik Data Traveller Pada Perusahaan Tour and Travel," *J. Manaj. Inform. dan Sist. Inf.*, vol. 2, no. 2, p. 61, 2019, doi: 10.36595/misi.v2i2.105.

11. Y. Sirisathitkul, P. Thanathamathee, and S. Aekwarangkoon, "Predictive apriori algorithm in youth suicide prevention by screening depressive symptoms from patient health questionnaire-9," *TEM J.*, vol. 8, no. 4, pp. 1449–1455, 2019, doi: 10.18421/TEM84-49.

12. H. Amin *et al.*, "Predictive Analysis of Heart Disease Using K-Means and Apriori Algorithms," *J. Appl. Sci. Comput.*, vol. VI, pp. 2183–2189, 2019, [Online]. Available: https://hal.archives-ouvertes.fr/hal-02267679

13. I. Pratama and P. T. Prasetyaningrum, "Pemetaan Profil Mahasiswa Untuk Peningkatan Strategi Promosi Perguruan Tinggi Menggunakan Predictive Apriori," *J. Eksplora Inform.*, vol. 10, no. 2, pp. 159–166, 2021, doi: 10.30864/eksplora.v10i2.505.

14. A. Kumar *et al.*, "A review of multi criteria decision making (MCDM) towards sustainable renewable energy development," *Renew. Sustain. Energy Rev.*, vol. 69, no. November 2016, pp. 596–609, 2017, doi: 10.1016/j.rser.2016.11.191.

15. F. M. Nafie Ali and A. A. Mohamed Hamed, "Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents," *J. Inf. Telecommun.*, vol. 2, no. 3, pp. 231–245, 2018, doi: 10.1080/24751839.2018.1448205.

16. D. Bansal and L. Bhambhu, "Usage of Apriori Algorithm of Data Mining as an Application to Grievous Crimes against Women," *Int. J. Comput. Trends Technol.*, vol. 4, no. 9, pp. 3194–3199, 2013, [Online]. Available: http://ijcttjournal.org/Volume4/issue-9/IJCTT-V4I9P144.pdf