International Journal of Computer Science and Software Engineering (IJCSSE), Volume 7, Issue 9, September 2018
ISSN (Online): 2409-4285 www.IJCSSE.org Page: 212-217

Visit Patterns Analysis of Foreign Tourist in Indonesian Territory Using Frequent Pattern Growth (FP-Growth) Algorithm

Emilia Regar[1] and Widya Silfianti[2]

[1] Business Information System, Gunadarma University, Jakarta Indonesia

[2] Computer Science, Gunadarma University, Jakarta Indonesia

[1]emiliaregar@gmail.com, [2]wsilfi@staff.gunadarma.ac.id

ABSTRACT

Tourism is a strategic sector because it encourages the creation of jobs, increasing public income, improving the quality of society and increasing foreign exchange. The government plans a tourist destination of 20 million foreign tourists coming to Indonesia in 2019. Indonesia's tourism marketing is not limited to increasing the visit of foreign tourists in the leading destinations but also increase the number of foreign tourists visiting the potential tourist region in Indonesia by looking at the pattern of foreign tourists visit. Data from the Ministry of tourism can be utilized by looking at spatial elements such as the map of the region is very strategic as the basic ingredients of policy development of tourist patterns for tourist routes or tourist recommendations. FP-Growth algorithm in data mining is applied as a technique to analyse the pattern of foreign tourists visiting Indonesia. This algorithm establishes the most frequent itemset of FP-Tree in a set of data. Data is processed using Weka software. Weka have an FP-Growth algorithm that result is tourist visits island of Sumatra, Java, Bali & Nusa Tenggara, Kalimantan, Sulawesi, Maluku and Papua and trans-island in Indonesia. Knowing the visit pattern of foreign tourists in Indonesia is expected to assist the relevant parties in the development of Indonesian tourism marketing.

Keywords: *Data Mining, Frequent Growth Pattern, Weka, Tourism, Destination, Foreign Tourist*.

1. INTRODUCTION

Tourism has become the fastest growing economic sector in the world. According to the World Tourism Organization (UNWTO), the number of international tourist arrivals in 2016 amounted to 1,235 billion or an increase of 3.9% compared to 2015 of 1.189 billion. UNWTO estimates that by 2030 international tourist arrivals will reach 1.8 billion. [1] UNWTO estimates are certainly seductive tourism business actors in various countries, then it takes innovation and appropriate strategies, and productive to seize the tourism market. [2]

Foreign tourists visiting Indonesia in 2016 increased by 15.54% ie 12,023,971 visits compared to the year 2015 10,406,759 visits, while foreign exchange earned by 12.44 billion USD or an increase of 1.75% over the year 2015 amounted to 12.23 billion USD. [1]

Tourism is very strategic sector because it becomes a sector that encourages the creation of jobs, increasing public income, improving the quality of society and increasing foreign exchange. The development of tourism in Indonesia needs to be directed so that tourism activities become one of the sectors that can drive other sectors of the economy. [1]

The government is planning a tourist destination of 20 million foreign tourists coming to Indonesia in 2019. In increasing foreign tourist arrivals, the government continues to develop tourism in Indonesia in terms of infrastructure, transportation and tourism marketing. The government's efforts in Indonesian tourism marketing are branding "Wonderful Indonesia" in information rooms and providing information in terms of cultural festival activities and tour packages for popular destinations through the website www.indonesia.travel. These efforts have an impact on the increase of foreign tourist visits along with the development of digital information, but the level of foreign tourists visit to Indonesia is still lower when compared with other countries in Asean such as Thailand and Singapore.

Seeing competitor countries whose visits are still higher government need the right strategy to increase the visits of foreign tourists to Indonesia. In addition to tourism management such as infrastructure and transport that need to be considered, Indonesia's tourism marketing is

an important factor in increasing the visits of foreign tourists. Indonesia's tourism marketing is not limited to increasing the number of foreign tourists visiting popular destinations but also increasing the number of tourists visiting Indonesia's potential destinations by looking at the pattern of foreign tourists. The pattern of visits of foreign tourists can be analysed from the data of the ministry of tourism. Utilizing data from the Ministry of tourism by looking at spatial elements such as the map of the region is very strategic as the basic of policy development of tourist patterns for tourist routes or tourist recommendations.

One technique that can be used to view visit patterns is data mining. Data mining helps to find characteristic patterns of data from a set of data. Data analysis using data mining to find the visit patterns of foreign tourists is an association analysis. Frequent Pattern Growth (FP-Growth) algorithm is an algorithm in association analysis. FP-Growth algorithm in data mining is applied as a technique to analyse the pattern of visits of foreign tourists visiting Indonesia to see potential destinations that can increase the visits of foreign tourists to Indonesia.

2. BASIC THEORY

Data mining is a combination of computer science disciplines that are defined as the process of discovering new patterns from massive data sets. Data mining is intended to extract (retrieve the essence) of knowledge from a set of data so that it can be understood. [3]

The importance of an associative rule can be determined by two parameters, support and confidence. Support is the percentage of combinations of items in the database and confidence is a strong inter-item relationship in association rules. The basic methodology of association analysis is divided into two stages frequent pattern mining and establishment of association rules. [4]

Patterns or rules that are formed from the value of support and confidence can be seen the size of interest or interestingness measures. Interestingness measures see whether the pattern or rule can be identified as something interesting or not even if for example A does not imply occurrence of incident B. In this case the value of the lift and leverage can be considered as alternative measures that give an indication of the correlation of occurrence A to B. Leverage, is parameter of a and b appearing not together in the dataset and it is expected that a and b are not independent of each other. Lift, the frequent a and b appears simultaneously [5]. The value of lift close to 1 means the antecedent and consequent of the rules or patterns are independent and the rules are not interesting if the value is higher means the rules is more interesting. [6]

FP-Growth is an alternative algorithm that can be used to determine the most frequent set of data sets in a set of data. [7] The data structure used to search frequent itemset with the FP-growth algorithm is an extension of the use of a prefix tree, commonly referred to as FP-tree. Using FP-tree, FP-growth algorithm can directly extract frequent itemset from FP-tree which has been formed by using divide and conquer principle. [8] FP-growth algorithm is divided into two main steps:

1. Build FP-tree. The support count of each item on each conditional pattern base is summed, then each item that has a larger number of support counts equal to the minimum support count will be generated by conditional FP-tree.
2. Find frequent itemset. If Conditional FP-tree is a single path, then a frequent itemset is obtained by combining items for each conditional FP-tree. If it is not a single track, then FP-growth is generated recursively. [8]

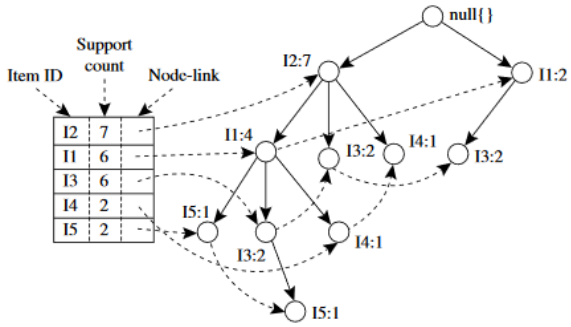

Fig. 1. FP-Tree to find Frequent Itemset [9]

In general, when considering the branch to be added for a transaction, the count of each node along a common prefix is incremented by 1, and nodes for the items following the prefix are created and linked accordingly. To facilitate tree traversal, an item header table is built so that each item points to its occurrences in the tree via a chain of node-links. The tree obtained after scanning all the transactions is shown in Figure 1 with the associated node-links. In this way, the problem of mining frequent patterns in databases is transformed into that of mining the FP-tree. [9]

3. RESEARCH METHOD

Flow analysis of the visit pattern begins with the selection of what dataset will be processed, then starting from preprocessing data. After the preprocessing phase is completed, next step is find the pattern using Weka. Weka have FP-Growth algorithm to produces the visit

pattern on foreign tourists. The flow of visits pattern analysis is shown in Figure 2.

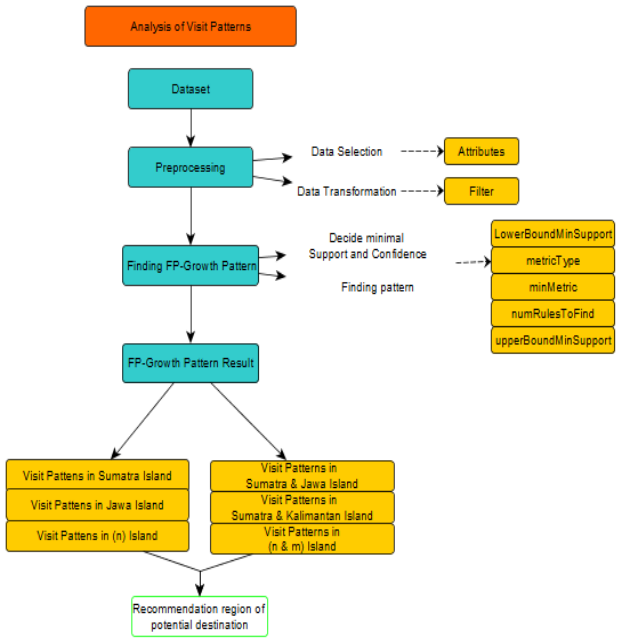

Fig. 2. Analysis of Visit Patterns

FP-Growth will generate tourist visit pattern which is grouped by 5 big islands covering Sumatra island, Java, Bali & Nusa Tenggara island, Kalimantan island, Sulawesi island and Maluku & Papua island. Travel patterns can also be trans-islands. An overview of the dataset processing flow on Weka to produce a tourist visit pattern shown in Figure 3.

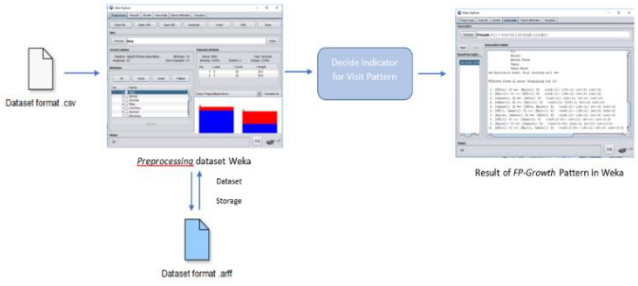

Fig. 3. Process to find the visit patterns of tourist on Weka

4. DISCUSSION

The dataset used to analyse the visit patterns of foreign tourists are dataset from 2014, 2015 and 2016 consisting of 34 Provinces visited with 53 countries of origin in 2016 and 51 countries of origin by 2015 and 2014. Examples of dataset shown are Sumatra island.

The pre-processing stage is doing data selection and data transformation. The pattern will display the name of region are possible to visit. The original data containing percentage of a foreign tourist changed to binary. "1" is indicate that the area was visited by a foreign tourist and "0" for the unvisited territory.

Find the visit pattern by FP-Growth method. The FP-Growth method uses FP-Tree as a frequent search of items in the database. The following will be described manual calculations in building FP-Tree so as to produce visit patterns of foreign tourist. After table 1 is changed to binary, Sort the territory or object with the most country visits on the number of items in the database. Examples of calculated data are Sumatra island data.

Table 1: Sequence of Territory

No.	Territory	Sum Countries of Origin
1	Kepualauan Riau	49
2	Sumatra Barat	48
3	Sumatra Utara	46
4	Nanggroe Aceh Darussalam	24
5	Riau	21
6	Bangka Belitung	20
7	Sumatra Selatan	18
8	Lampung	13
9	Jambi	11
10	Bengkulu	6

From table 1 create the FP-Tree based on the highest number of visits to the lowest. The minimum value of support for Sumatra province is set at 30%. 30% means Territories with less than 16 of the total number of 53 countries visiting Indonesia are not included as search object visits. The area of Lampung, Jambi and Bengkulu does not meet the minimum support limit then the region does not need to be included in the FP-Tree node. FP-Tree mapping to find the pattern of visits starting from the bottom to the top node. For initial mapping starts from looking for patterns ending in node 7 or Sumsel region. FP-Tree in figure 4 that ends with node 7. The initial stage is searched for a visit pattern with 2 items. if still meet the support limit then continue looking for the pattern of visits 3 items up to 4 items. The results are shown in Table 2.

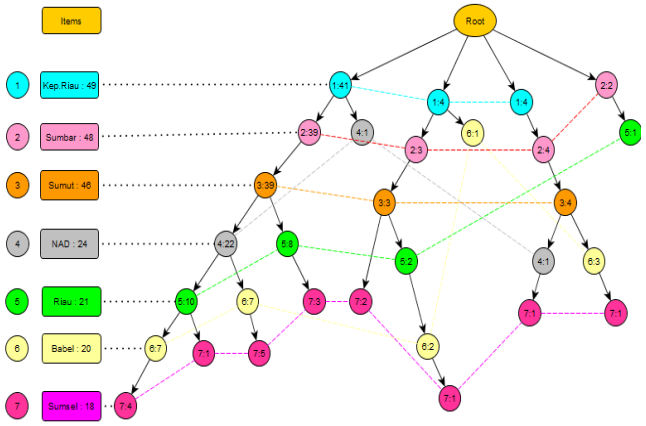

Fig. 4. FP-Tree Sumatra Island

Table 2: Combination of Visit Patterns ends with Sumsel

Combination of Visit Pattern 2 Items						
No	**Antecedent**	**Consequent**	**Supp (%)**	**Conf (%)**	**Lift**	**Lev**
1	Sumut	Sumsel	33,96	39,13	1.15	0.03
2	Sumsel	Sumut	33,96	100	1.15	0.04
3	Sumsel	Sumbar	33,96	100	1.1	0.04
4	Sumbar	Sumsel	33,96	37,5	1.1	0.03
5	Sumsel	Kep.Riau	33,96	100	1.08	0.03
6	Kep.Riau	Sumsel	33,96	36,73	1.08	0.03
Combination of Visit Pattern 3 Items						
No	**Antecedent**	**Consequent**	**Supp (%)**	**Conf (%)**	**Lift**	**Lev**
1	Sumbar-Sumut	Sumsel	33,96	39,1	1,15	0,04
2	Sumsel	Sumbar-Sumut	33,96	100	1,15	0,04
3	Sumut-Sumsel	Sumbar	33,96	100	1,1	0,03
4	Sumbar	Sumut-Sumsel	33,96	37,5	1,1	0,03
5	Sumbar-Sumsel	Sumut	33,96	100	1,15	0,04
6	Sumut	Sumbar-Sumsel	33,96	39,1	1,15	0,04
Combination of Visit Pattern 4 Items						
No	**Antecedent**	**Consequent**	**Supp (%)**	**Conf (%)**	**Lift**	**Lev**
1	Sumsel	Kep.Riau-Sumbar-Sumut	33,96	100	1,15	0,04
2	Kep.Riau-Sumbar-Sumut	Sumsel	33,96	36,7	1,15	0,04
3	Sumut	Kep.Riau-Sumbar-Sumsel	33,96	39,1	1,15	0,04
4	Kep.Riau-Sumbar-Sumsel	Sumut	33,96	100	1,15	0,04
5	Sumbar	Kep.Riau-Sumut-Sumsel	33,96	39,1	1,15	0,04
6	Kep.Riau-Sumut-Sumsel	Sumbar	33,96	100	1,15	0,04
7	Kep.Riau	Sumbar-Sumut-Sumsel	33,96	36,7	1.08	0.03
8	Sumbar-Sumut-Sumsel	Kep.Riau	33,96	100	1.08	0.03

Information
Supp = Support
Conf = Confidence
Lev = Leverage

Taking a sample pattern ending in South Sumatra (Sumsel) in table 2 to outline the support, confidence, lift and leverage calculation from FP-Tree. Example calculation for [Sumut-Sumsel → Sumbar] pattern

- Support (A,B) =

$$P(A \cap B) = \frac{\text{Sum of visitors in territory A and B}}{\text{Total sum of country visits}} \quad (1)$$

Support Sumut-Sumsel-Sumbar

$$= \frac{\text{Sum of visitors Sumut}-\text{Sumsel}-\text{Sumbar}}{\text{Total sum of country visits}}$$

$$= \frac{18}{53} = 0{,}339 \text{ or in percent } 33{,}9\%$$

- Confidence (A) =

$$P(B|A) = \frac{\text{Sum of visitors in territory A and B}}{\text{Sum of visitors in territory A}} \quad (2)$$

Confidence [Sumut-Sumsel → Sumbar]

$$= \frac{\text{sum of visitors Sumut}-\text{Sumsel}-\text{Sumbar}}{\text{sum of visitors Sumbar}}$$

$$= \frac{18}{18} = 1 \text{ or in percent } 100\%$$

- Lift = $\frac{\text{Support}(A \cap B)}{(\text{Support}(A) \text{ x } \text{Support}(B))}$ (3)

$$\text{Lift} = \frac{\text{Support (Sumut}-\text{Sumsel)ke Sumbar}}{\text{Support Sumut}-\text{Sumsel x Support Sumbar}}$$

$$= \frac{0.339}{(0.339 \text{ x } 0.867)} = 1.15$$

- Leverage = P(A∩B) – (P(A) x P(B)) (4)

Leverage = Support (Sumut-Sumsel-Sumbar) – (Support (Sumut-Sumsel) x Support(Sumbar)
= 0.339 – (0.339 x 0.867) = 0.04

The value of lift 1.15 and leverage 0.04 on the visit pattern [Sumut-Sumsel → Sumbar] with a value of 33.96% support and 100% confidence means that the pattern of the visit is interesting with items that are dominantly independent because the leverage value approaches 0. The FP-Growth method is also done to find the pattern of visits using Weka on the islands of DKI Jakarta, Bali & Nusa Tenggara, Kalimantan, Sulawesi, Maluku and Papua but the results shown only Sumatra island can be seen in table 3.

Table 3: Visit Patterns of Sumatra Island processed by Weka

No	Antecedent	Consequent	Supp	Conf	Lift	Lev	Year
1	Sumbar	Sumut	0.47	1	1.34	0.12	2014
2	Sumsel	Sumut	0.33	1	1.34	0.08	2014
3	Kep.Riau,Sumbar	Sumut	0.43	1	1.34	0.11	2014
4	Kep.Riau,Sumsel	Sumut	0.31	1	1.34	0.08	2014
5	Sumsel	Kep.Riau	0.31	0.94	1.37	0.08	2014
6	Sumut,Sumsel	Kep.Riau	0.31	0.94	1.37	0.08	2014
7	Sumbar	Kep.Riau	0.43	0.92	1.34	0.11	2014
8	Sumut,Sumbar	Kep.Riau	0.43	0.92	1.34	0.11	2014
9	Kep.Riau,Sumut	Sumbar	0.43	0.73	1.56	0.15	2014
10	Kep.Riau,Sumbar	Sumut	0.49	0.93	1.39	0.14	2015
11	Sumbar	Sumut	0.55	0.9	1.35	0.14	2015
12	Kep.Riau,Sumut	Sumbar	0.49	0.89	1.47	0.16	2015
13	Sumut,Sumbar	Kep.Riau	0.49	0.89	1.3	0.11	2015
14	Sumbar	Kep.Riau	0.53	0.87	1.27	0.11	2015
15	Sumut	Kep.Riau	0.55	0.82	1.2	0.09	2015
16	Sumut	Sumbar	0.55	0.82	1.35	0.14	2015
17	Kep.Riau	Sumut	0.55	0.8	1.2	0.09	2015
18	Kep.Riau	Sumbar	0.53	0.77	1.27	0.11	2015
19	Sumut	Sumbar	0.87	1	1.1	0.08	2016
20	Kep.Riau,Sumbar	Sumut	0.87	1	1.15	0.11	2016
21	Kep.Riau,Sumut	Sumbar	0.87	1	1.1	0.08	2016
22	Sumbar,NAD	Sumut	0.43	1	1.15	0.06	2016
23	Sumut,NAD	Sumbar	0.43	1	1.1	0.04	2016
24	Sumbar,Sumsel	Sumut	0.34	1	1.15	0.04	2016
25	Kep.Riau,Sumbar,NAD	Sumut	0.44	1	1.15	0.06	2016
26	Kep.Riau,Sumut,NAD	Sumbar	0.43	1	1.1	0.04	2016
27	Sumbar	Sumut	0.87	0.96	1.1	0.08	2016
28	Kep.Riau,NAD	Sumut	0.44	0.96	1.1	0.04	2016

An interesting pattern of visits from the table 3 is the pattern of visits to 9, 12 and 21 [Kep.Riau, Sumut → Sumbar]. The pattern of this visit gets the value of support and confidence that increases each year with a pattern that has the independence of items increases each year due to the value of lifts and leverage are declining.

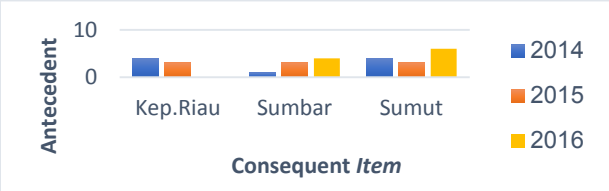

Fig. 5. Graph of potential tourist areas of Sumatra island

Figure 5 of potential areas to be visited on the island of Sumatra in 2014 and 2015 are Riau Islands (Kep.Riau), West Sumatra (Sumbar) and North Sumatra (Sumut) while in 2016 the province of Sumut and Sumbar. These three regions arise from the combination of visits patterns generated in table 3 Overall the area on the island of Sumatra is always visited by tourists are Kep.Riau, Sumbar and Sumut. The areas that also appear on the pattern of visits are Nanggroe Aceh Darusallam (NAD) in 2016 and South Sumatra (Sumsel) in 2014.

5. CONCLUSIONS

The process of searching the visit pattern of foreign tourists based on FP-Growth algorithm give result the visit pattern of foreign tourist visits in the territory of Indonesia. FP-Growth also gives recommendation of potential destination by looking at the pattern of foreign tourist visits in Indonesia from consequent item on visit pattern. Consequent items become potential areas because consequent items is an effect after another destination has visited.

The results from the visit pattern of foreign tourist in Indonesia found some of the same pattern in every year and there are also patterns that exist only in a certain year. The visit pattern that changes in every year indicate an increase in visits in each region and the areas become an independent item (not dependent on another items) while for the same pattern of visits in some areas indicates that the region visited occurs in one area only so the local govement need to support the area. Items on the visit pattern are also influenced by popular destination such as province in Bali, DKI Jakarta, Riau Islands, DI Yogyakarta, West Java, Central Java, East Java, West Nusa Tenggara, Papua and West Sumatra. It expected that the potential destinations increase tourist visits to Indonesia in addition to popular destination along with the development of tourism in 10 priority destinations of Lake Toba in North Sumatra, Kepulauan Seribu in DKI Jakarta, Tanjung Lesung in Banten, Borobudur in Central Java, Bromo Tengger Semeru in East Java, Labuan Bajo in NTT, Tanjung Kelayang in Bangka Belitung, Mandalika in NTB, Wakatobi in Southeast Sulawesi and Morotai in North Maluku. Tourism marketing strategy by combining routes on popular and potential destinations is expected to increase tourist visits in Indonesia.

Proposed for further research the existing data can be studied down to the level of region in Indonesia that can bring up a potential destination. The results expected to provide information about tourism development strategy in Indonesia apart from the strategies that have been implemented by the government.

REFERENCES

[1] Kemenpar, 2016. Statistik Profil Wisatawan Mancanegara 2016. [Online] Available at: http://www.kemenpar.go.id [Accessed 28 Mei 2017].
[2] Kemenpar, 2016. NESPARNAS Neraca Satelit Pariwisata Nasional 2016. [Online] Available at: http://www.kemenpar.go.id [Accessed 26 Juli 2017].
[3] Suyanto, 2017. Data Mining Untuk Klasifikasi Dan Klasterisasi Data. Bandung: Informatika.
[4] Kusrini & Luthfi, E. T., 2009. Algoritma Data Mining. Yogyakarta: ANDI.
[5] Maimon, O. & Rokach, L., 2005. Data Mining and Knowledge Discovery Handbook. 2nd ed. New York: Springer.
[6] Kotu, V. & Deshpande, B., 2015. Predictive Analytics and Data Mining: Concepts and Practice with Rapid Miner. Waltham USA: Morgan Kaufmann.
[7] Ikhwan, A., Nofriansyah, D. & S., 2015. Penerapan Data Mining dengan Algoritma FP-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma). Journal of SAINTIKOM, Volume Vol.14.
[8] Gunadi, G. & Sensuse, D. I., 2012. Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth) : Studi Kasus Percetakan PT.Gramedia. Journal of TELEMATIKA MKOM, Volume Vol.4.
[9] Han, J., Kamber, M. & Pei, J., 2012. Data Mining Concepts and Techniques. 3rd ed. Waltham: Morgan Kaufmann.
[10] Bramer, M., 2016. Principles of Data Mining. 3rd ed. Verlag London: Springer.
[11] Fakih, M. F., 2017. Penentuan Pola Kunjungan Wisatawan Ke Berbagai Obyek Daya Tarik WIsata di Pulau Ambon menggunakan Metode Frequent Pattern Growth. Institut Teknologi Sepuluh November, Program Magister Telematika, Fakultas Teknologi Industri.
[12] Han, J., Kamber, M. & Pei, J., 2006. Data Mining Concept and Techniques. 2nd ed. Waltham: Morgan Kaufman Publishers.
[13] Law of R.I. No.10 / 2009 on tourism
[14] Nofriansyah, D., 2014. Konsep Data Mining vs Sistem Pendukung Keputusan. Yogyakarta: Deepublish.
[15] Prasetyo, E., 2014. Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: ANDI.
[16] Purnamasari, D. et al., 2013. Get Easy Using Weka. Jakarta: Dapur Buku.
[17] Soman, K. P., Diwakar, S. & Ajay, V., 2014. Insight Into Data Mining: Theory and Practice. Delhi: PHI Learning Private Limited.