Tourism Trend Mapping Based on Social Media Using SAW Algorithm

M A Maulana¹, M habib², A Setyanto³ and Oktavia⁴

¹,²,³ Magister of Information Engineering, University of Amikom Yogyakarta, Indonesia.
⁴ Bachelor of Information Engineering, University of Amikom Yogyakarta, Indonesia.

E-mail: muhammad.0874@students.amikom.ac.id

Abstract. Indonesia is a country that has a variety of places to hold holidays. Having a plan about the most popular tourist attractions is the key for everyone who wants to travel to an area. But data on tourist spots that are currently in trend are often not easily obtained. Different from tour agents who already have a lot of experience and data without fear of missing information, therefore, presenting the tourism trends in a particular group is almost impossible to use data on the number of frequented places. This paper will discuss tourism trends that have appeal based on social media, with case studies on Twitter data with certain regional fields. Social media provides large raw data including travelers' choices when they visit the venue. However, the data obtained is not shaped in human language. These results will be processed first and will be arranged to determine the level of tourist destinations are often visited. The algorithm used to process using Simple Additive Weighting (SAW) is proposed to arrange food ranking. To measure accuracy, we compared our results with tourist sights in Yogyakarta.

1. Introduction

Traveling has become a lifestyle that can’t be separated from everyday life because holiday is the main requirement of all people. Tourism is also one form of identity and expressive cultural heritage. Tourism is also being used as a commodity for a culture-based creative industry. This is because every region in Indonesia has a unique tourism potential and various types. Sourced some statistical data in the 2017 range, the tourism sector contributes about 20% of foreign exchange contributors in Indonesia [1]. Places where migrants have a high chance of providing jobs for the community. Indonesia is a strategic sector in the development of the tourism industry including Yogyakarta. Yogyakarta is famous for tours such as Malioboro, Borobudur Temple, Alun-alun Kidul (Alkid) and so on. In addition, for local governments, tourist data becomes very important to be used as a reference supporters of tourism development. Periodic tourist data cannot be readily available to the public. This is because important data is a tourism development strategy in creating opportunities [2]. Therefore, social media is one way find out tourism opportunities. On the other hand, the use of social media can be used as a reference for tourism development needs that can be used as a means of promotion in getting closer to tourists. [3]. Twitter is one of the social media platforms that is used to promote tourism through posts, comments or tweets about tours, Twitter and Facebook account users can comment on tours that have their own uniqueness [4]. This is used to find trends in tourist attractions based on the most choices of public who have visited certain places. This research can be used to...
create tourism development opportunities by knowing the recommendation of tourist trends. In addition, for travelers, visiting a city in Indonesia is required to explore the typical tourist areas that are trend [5]. Therefore, the importance of a system that can recommend tourist attractions which is a trend is based on tweets, posts and comments from social media when making tourist visits in a particular area.

The algorithm applied to classify tourism data related to Twitter is by using the Simple Additive Weighting (SAW) algorithm to search for tourism groups that are the topic of conversation on social media. Then to determine the level value with the method commonly used is Simple Additive Weighting (SAW). Research related to exploring opportunities to use data from social media platforms to predict post-conditions, however, they do not work on particular topics of interest and lack of evaluation criteria [6]. This study explores the possibility of using social media to predict tourist trends.

2. Related Work

Research on the building area, wide parking, accessibility, safe distance to the city center, location and convenience prices can increase sales for the tourism sector and culinary tourism sector. This study used one of the main methods by searching for the weighted performance quantities of each alternative with simple additive weighting method (SAW) is looking for a weighted sum of the performance of each alternative. The studies produce rankings to determine new places [7]. Research that positions singers play an important role in the band to be able to influence the audience. This study uses two methods to find the best singer with analytic healing process (AHP) and priority level using SAW. This method can obtain consistency accuracy of 84.61% [8]. Forest monitoring research is one way to maintain the preservation of urban forests. This study uses one method by the decision makers of the analysis of urban forest health monitoring with Simple Additive Weighting (SAW) [9].

3. Methodology

This research is a research conducted by observation, data taken from social media and travel agents. This research was conducted with 6 stages, namely data collection, data cleaning, feature extraction, grouping, ranking, and evaluation. The output of this study was to compare the accuracy of tourism trend predictions with data from travel agents.

The results of this study are based on web-based social media trends. In general, an overview of the steps in this study in Figure. 1.
3.1 Crawling data
The first step starts with retrieving data on social media, in this study data was obtained from Twitter, for data retrieval by utilizing the Twitter Application Programming Interface (API). Data crawled with tourism topics, which are in the city of Yogyakarta. Data in the form of tweets or comments from users who provide opinions about the tour you have visited. In this step the thing that needs to be considered is the time used as a tourist trend.

3.2 Cleaning
After getting data from Twitter social media, then the data is cleaned up. because the data obtained from Twitter is still a lot of noise that is not needed then data cleaning is done to differentiate data sources with different characteristics from one data to another. In this step, deletion of symbols and filtering of unnecessary information is needed to enter the next stage. Data that has been cleared, will be converted in the pocket of the word matrix to apply the machine learning algorithm. further feature extraction will be carried out to find features that are suitable for the trendy tourist data classification.

3.3 Ranking
The next step is a Simple Additive Additive algorithm which is used to classify the type of tour data based on the group that has been determined. Simple Additive Weighting aims to divide n and observe in group k where each inhibition is in a cluster with the closest mean [14]. This results in partitioning the data space into Voronoi cells. Given a set of observations (x1, x2, ..., xn), where each observation is a real d-dimensional vector, Simple Additive Weighting aims to separate n observations into groups k G = {G1, G2, Gk} so for reduce the number of squares in a cluster defined as follows:

$$r_y = \begin{cases} \frac{x_{y}}{\text{Max} x_{y}} \\ \frac{x_{y}}{\text{Min} x_{y}} \end{cases}$$

Data ranking is done using SAW method. the initial baseline of the SAW method is to determine the number of performance ranking groups in each alternative from the existing attributes [16]. The SAW
method requires the decision matrix normalization process \( X \) to a proportional scale for all alternative ratings. Given the following equation: Where \( r_{ij} \) is the normalized performance rank of the alternate \( A_i \) in the attribute \( C_j \); \( i = 1,2, \ldots, m \) and \( j = 1,2, \ldots, n \). \( V_i \) means the value of preference, such as weighting and \( r_{ij} \) as a normalized performance level. The preference value for each alternative \( (V_i) \) is given as follows:

\[
V_i = \sum_{j=1}^{n} W_j r_{ij}
\]

(2)

A larger \( V_i \) value indicates that \( A_i \)'s alternative is more dominant. The SAW method has several stages, namely to determine the type to be used as a reference in decision making, namely \( C \). Then, determine the ranking of each alternative in each criterion then make a decision matrix based on criteria \( (C) \), to obtain normalized matrix \( R \) then normalize Matrices are based on equations that are adjusted for attribute types (gain attributes or cost attributes). The final result obtained from the ranking process is the number of multiplication of the normalization matrix \( R \) with the weight vector to get the largest value chosen as the best alternative \( (A) \) as the solution. There are several advantages of the SAW algorithm compared to other algorithms for decision making, which lies in its ability to make judgments that are more appropriate because they are based on predetermined preference values and criteria. not only that, SAW can also be an alternative, from several alternatives because of the ranking grouping process after determining the weight value of the existing attributes.

4. Result and Discussion

Data from Twitter collected for three weeks as a sample. Tweets obtained from the Yogyakarta place criteria are divided into 3 keywords:

1. Nature : pine forest, gembira loka zoo, etc.
2. History : ratu boko temple, prambanan temple garden, etc.
3. Beach : indrayanti beach, barong beach, etc.

Thoses are downloaded through twitter API with unstructured text. From the survey results where collected from several travel agents who often bring tourists around Yogyakarta. The result of location recommendation in Yogyakarta are shown in Fig 2 below, from this algorithm can make it easier for travelers to determine what tourist destinations are trending.

Figure 2. Data prediction travel agent
4.1 Calculation of SAW

Based on the chart shows nature has the highest trend for tourist location trends. Then, the graph obtained can be explained in detail and steps of the following formula. The criteria used in the recommendation of tourism location decision are 2 criteria. The 2 criteria used can be seen in Table I.

| Code | Criteria      | Weight |
|------|---------------|--------|
| C1   | Posting Time  | 0.25   |
| C2   | Emergen       | 0.55   |

After determining the criteria and weight for the dataset, then specify the standard value of the SAW algorithm as explained repeatedly:

1) **Very Rare (VR)** = 0
2) **Rare (R)** = 2
3) **Moderate (M)** = 3
4) **High (H)** = 4
5) **Very High (VH)** = 5

Twitter data obtained from extraction of features that have been converted based on alternatives and criteria can be written as:

1) **Nature** = A1
2) **History** = A2
3) **Beaches** = A3

4.2 Alternative Value Criteria

Alternative values in each criterion can be explained in Table II.

| Code | Result |
|------|--------|
| A1   | 5 5    |
| A2   | 3 2    |
| A3   | 2 1    |

After obtaining alternative weight that has been adjusted to the appropriate value, then enter the normalization stage in formula (2).

Alternative-1 (A1):

$$r_{11} = \frac{5}{(5+5)} = 0.5$$
\[ r_{1.2} = \frac{5}{(5+5)} = 0.5 \]

Alternative-2 (A2) :
\[ r_{2.2} = \frac{3}{(3+2)} = 0.6 \]
\[ r_{2.2} = \frac{2}{(3+2)} = 0.4 \]

Alternative-3 (A3) :
\[ r_{3.2} = \frac{2}{(2+1)} = 0.6666666667 \]
\[ r_{3.2} = \frac{1}{(2+1)} = 0.3333333333 \]

4.3 Normalized
Normalization process is carried out until alternative calculations up to three and the normalization results obtained are shown in Table 3.

| Code | Result |
|------|--------|
|      | C1     | C2     |
| A1   | 0.5    | 0.5    |
| A2   | 0.6    | 0.4    |
| A3   | 0.6666666667 | 0.3333333333 |

After the results are obtained from normalization, to get the rating of all alternatives, a multiplication matrix (preference) is carried out. The weighted values given can be explained in Table IV below.

| Code | Result |
|------|--------|
| A1   | \((0.25 \times 0.5) + (0.55 \times 0.5)\) |
| A2   | 0.37   |
| A3   | 0.35   |

From this result, we can draw the conclusion that the full marks is: natural code (A1) with a value of 0.38846153845. Thus, A1 can be concluded as a recommended tourist location for 3 weeks based on the SAW algorithm.

4.4 Evaluation
The evaluation is compared with travel agent data in several tourist place in Yogyakarta using SAW algorithm.
Table 5. Example data from social media

| Parangteritis panasnya tapi seru main air sambil makan ikan di pinggir pantai |
| Lava bantal guys? Dah pada kesana? Seru lo |
| Rafting di dket borobudur pas airnya tinggi banget lagi, untung enggak ada yg luka |
| Pilih candi mana ya di jogja? Prambanan asik nih kliatannya |

Table 6. Travel agent data

| Location | Week I | Week II | Week III |
|----------|--------|---------|----------|
| Nature   | 70     | 50      | 80       |
| History  | 50     | 50      | 60       |
| Beaches  | 40     | 40      | 50       |

from the two tables above (Table IV and Table VI) we can see the relationship between data on travel agents and twitter data, but it still has to be tested again for percentage accuracy of both data. With the results obtained, this shows that the algorithm applied in this paper can be used to search for trends in trends based on social media and this includes the analysis of big data for public users.

Acknowledgments
AMIKOM University Yogyakarta has provided support by providing facilities needed for research. This study can also be done on the good cooperation of all sectors, one of them is lecturer of big data subject Mr. Arief Setyanto

5. References

[1] E Simorangkir 2018 Pariwisata Jadi Andalan Penyumbang Devisa US$ 20 Miliar (Finance detik)
[2] M D Ramadhani, C Rahmad and F Rahutomo 2017 Tourism destination rating system based on social media analysis (proposal and dataset development in Indonesian language) (International Conference on Sustainable Information Engineering and Technology (SIET)) pp 41-46
[3] I Perikos and I Hatzilygeroudis 2018 A Framework for Analyzing Big Social Data and Modelling Emotions in Social Media (IEEE Fourth International Conference on Big Data Computing Service and Applications) Pp 80-84.
[4] N Blagus and S Žitnik 2018 Social media comparison and analysis: The best data source for research?, (12th International Conference on Research Challenges in Information Science (RCIS)) pp 1-10
[5] M M Mostafa 2013 More than words: Social networks’ text mining for consumer brand sentiments (Expert Syst Appl) vol 40 no 10 pp 4241–4251
[6] D Herumurti, R R Hariadi, I Kuswarday, Y Anny, N Suciati and S Arifiani 2017 iARTour for Indonesia tourism object (International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation (ICAMIMIA)) pp 165 – 170.
[7] A Hidayat, M Muslihudin and I T Utami 2017 Sistem Pendukung Keputusan Menentukan Lokasi Cafe Baru Suncafe Sebagai Destinasi Wisata Kuliner Di Kabupaten Pringsewu Menggunakan Metode Simple Additive Weighting (SAW) (Jurnal TAM (Technology Acceptance Model)) 6 pp 71-79

[8] A Cahyapratama and R Sarno 2018 Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) Methods In Singer Selection Process (International Conference on Information and Communications Technology (ICOIACT)) pp 234 – 239.

[9] A Pranolo and S M Widayastuti 2014 Simple additive weighting method on intelligent agent for urban forest health monitoring (Proceeding - 2014 Int Conf Comput Control Informatics Its Appl “New Challenges Oppor Big Data”)