Data mining techniques are being used to map the number of foreign guests at Indonesian star hotels

Debi Masri[1], Sri Kasnelly[2]

[1]*Universitas Battuta, Medan, North Sumatra, Indonesia*
[2]*STAI An-Nadwah Kuala Tungkal, Jambi, Indonesia*

[*]*debi_director@yahoo.com*

Abstract

Statistics on foreign tourist arrivals show an upward trend in recent years. This is beneficial because foreign exchange in the tourism sector is generated by spending by foreign tourists visiting Indonesia. Mapping the number of tourists who visit Indonesia is necessary to determine the extent of a region's natural potential. The number of foreign guests at Indonesia's star hotels is one indicator that can be used. The goal of this study is to use data mining techniques to visualize the mapping pattern of the number of foreign guests at five-star hotels in order to obtain important information. The Central Statistics Agency (abbreviated BPS) provided the data for the number of foreign guests at five-star hotels in 2017-2019. Clustering with the k-medoids technique is the data mining method employed. The Davies-bouldin index (DBI) operator is used to calculate the number of clusters. The number of good clusters was 3 (k=3), with cluster 0 (low) = 20 provinces, cluster 1 (normal) = 23, and cluster 2 (high) = 1 province, and a DBI value of 0.215 (k=3). According to the study's findings, the province of Bali is one of the areas with the greatest potential for foreign tourist visits (high cluster). Meanwhile, there are 20 provinces with a low number of foreign visitor visits, so this can be used to assess how to increase the number of foreign visitor visits to the province.

Keywords: *Data mining, number of foreign guests, hotels, k-medoid, mapping, Indonesia.*

1. Introduction

Indonesia has a lot of tourism potential. Tourism activities are expected to be a reliable development force. According to Tourism Law No. 10 of 2009, tourism aims to increase economic growth, improve people's welfare, eliminate poverty, overcome unemployment, promote culture, and improve the nation's image [1]. Because of its ability to bring in foreign exchange, the tourism industry is an invisible export. Foreign exchange in the tourism sector is generated by spending by foreign tourists visiting Indonesia. Statistics on foreign tourist visits show an increasing trend over the last few years, with the highest number of visits in 2019 [2].The increase in the number of tourists must be balanced by an increase in the number of available hotel rooms. Tourists who visit tourist destinations will almost certainly require a place to stay. As a result, the hotel serves as one of the benchmarks for determining how many tourists visit a particular area. The length of stay of hotel guests is an important indicator of how interested tourists are in the area visited. The longer tourists stay in a location, the more popular it is with tourists. The hotel room occupancy rate is also an important indicator of the hotel's success in selling rooms. The higher the occupancy rate, the more rooms that can be sold. The number of rooms used or sold each night is one indicator of hotel productivity.

In 2019, the Central Statistics Agency (abbreviated BPS) publishes statistical data on hotels and other forms of lodging in Indonesia. This information was released in early July 2020. The information presented was gathered from an annual survey of all companies, hotel businesses, lodging, and others conducted throughout the month. From April to November of 2019, Based on the presented data, namely the number of foreign
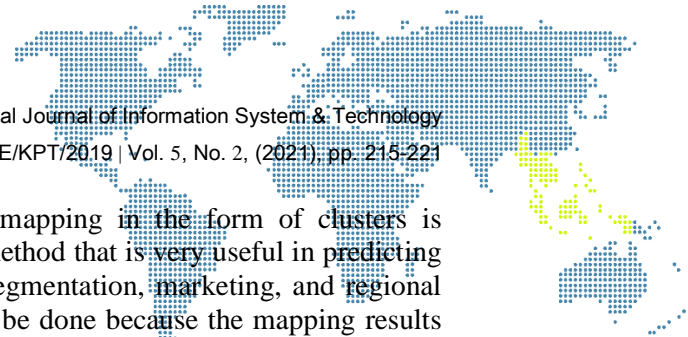
guests staying at five-star hotels by province, a mapping in the form of clusters is required, where this cluster is a data segmentation method that is very useful in predicting and analyzing certain business problems. Market segmentation, marketing, and regional zoning mapping are a few examples [3]. This must be done because the mapping results will be used as evaluation material by the central and local governments when making the next decision. A technique is required to visualize the data and extract some important information from the data so that it can be used to do useful work in order to see the mapping pattern of the number of foreign guests in existing star hotels []. K-medoids is one of the data mining methods [4]–[9]. This method will be used to categorize data on the number of foreign guests staying in star hotels in each province. The k-medoids method is a simple to implement algorithm that has a relatively fast performance, is easy to adapt, and is widely used [10], [11].

2. Research Methodology

For the number of foreign guests at star hotels in Indonesia, this study employs the k-medoids method in a clustering system. The Central Statistics Agency (abbreviated BPS) provided the dataset for the number of foreign guests at five-star hotels in 2017-2019. The mapping process is carried out by determining the number of clusters using the k-medoids method in conjunction with the Davies-bouldin index (DBI) method, which is one of the internal evaluation methods that measures the evaluation of the formed cluster [12], [13]. The number of the best clusters is indicated by the smaller DBI value [14]. The research flow for the k-medoid method is as follows:
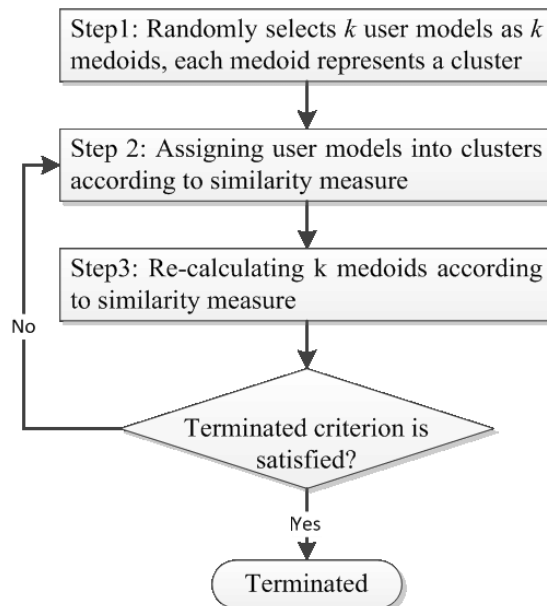

Figure 1. The research flow uses K-Medoids

The stages of this research are as follows:
a) Input dataset;
b) Set up the cluster center;
c) Distance calculation;
d) Choose new medoids at random;
e) Distance calculation;
f) Determine the total deviation (S);
g) Repeat steps 3–5 until no medoid change occurs;
h) Output in the form of clusters comprised of clusters determined based on the best DBI assessment results.

3. Results and Discussion

The dataset on the number of foreign guests at star hotels is presented in Table 1 in this study, and the analysis process employs the RapiMiner software.

Table 1. Dataset of the number of foreign guests at star hotels in 2017-2019

Province	Number of Foreign Guests at Star Hotels (Thousand People)		
	2017	2018	2019
Aceh	24.59	11.73	9.21
North Sumatra	314.78	346.94	201.88
West Sumatra	75.86	43.08	51.83
Riau	35.19	30.54	35.33
Jambi	7.89	5.66	6.02
South Sumatra	14.27	16.96	14.05
Bengkulu	1.07	0.70	0.56
Lampung	9.00	3.25	3.01
Kep. Bangka Belitung	13.17	8.18	9.68
Kep. Riau	1631.31	2497.80	1595.59
DKI Jakarta	1879.44	2266.08	1529.76
West Java	516.94	536.60	502.72
Central Java	140.70	112.21	122.15
DI Yogyakarta	317.93	237.59	211.50
East Java	287.27	238.70	267.97
Banten	437.11	496.89	464.21
Bali	5973.98	9252.94	5687.80
West Nusa Tenggara	317.10	211.60	195.80
East Nusa Tenggara	49.79	34.78	58.59
West Kalimantan	35.56	33.33	33.00
Central Kalimantan	7.69	5.47	6.68
South Kalimantan	13.10	12.10	9.41
East Kalimantan	37.62	23.41	30.66
North Kalimantan	2.86	2.44	2.57
North Sulawesi	128.53	148.36	164.71
Central Sulawesi	3.86	4.07	2.00
South Sulawesi	109.07	78.22	46.91
Southeast Sulawesi	3.21	2.36	1.40
Gorontalo	5.24	2.94	1.99
West Sulawesi	0.87	0.19	0.38
Maluku	6.45	10.68	7.57
North Maluku	0.53	0.63	0.66
West Papua	7.38	12.60	19.25
Papua	10.89	11.16	12.58

Source: Monthly Hotel Survey (VHTS), BPS
Source Url: https://www.bps.go.id/indicator/16/310/1/sum-tamu-asing-pada-hotel-star.html

In building the system it is necessary to make a design. The design of this system is intended so that in the formation of the system a system can be produced that functions optimally and provides convenience for its users. The analysis of the system to be built will be explained in detail in the following flowmap in figure 3.
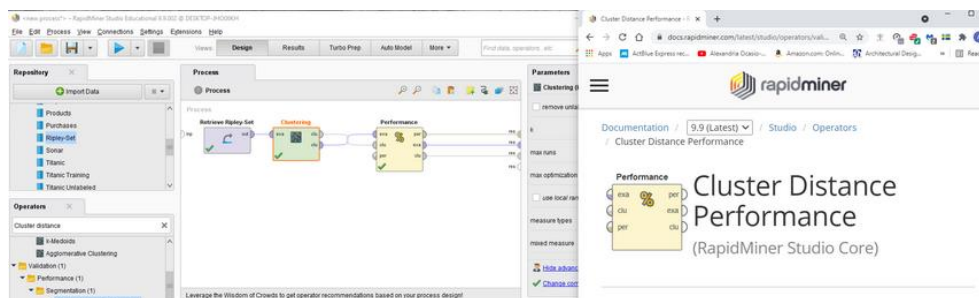

Figure 2. The design of k-medoids in determining the value of k
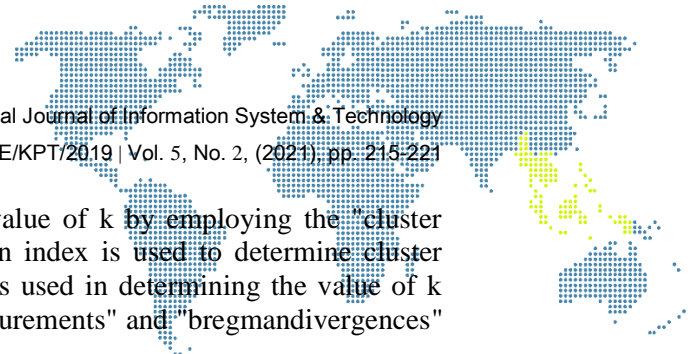
Figure 2 depicts the model used to calculate the value of k by employing the "cluster distance performance" operator. The davies-bouldin index is used to determine cluster evaluation with this operator (DBI). The parameters used in determining the value of k use two types of measures, namely "numerical measurements" and "bregmandivergences" as shown in Figure 3 below.
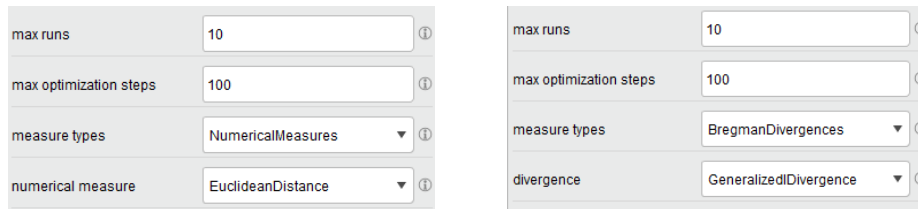

Figure 3. Comparative parameters in determining the value of k

The results of tests using two types of measures yield different k values, where DBI-1 represents the result of measures = "numerical measurements" and DBI-2 represents the result of measures = "bregmandivergences." The DBI results for the values of k=2, 3, 4, 5, and 6 are shown in Table 2 and Figure 4 below:

Table 2. Value from DBI

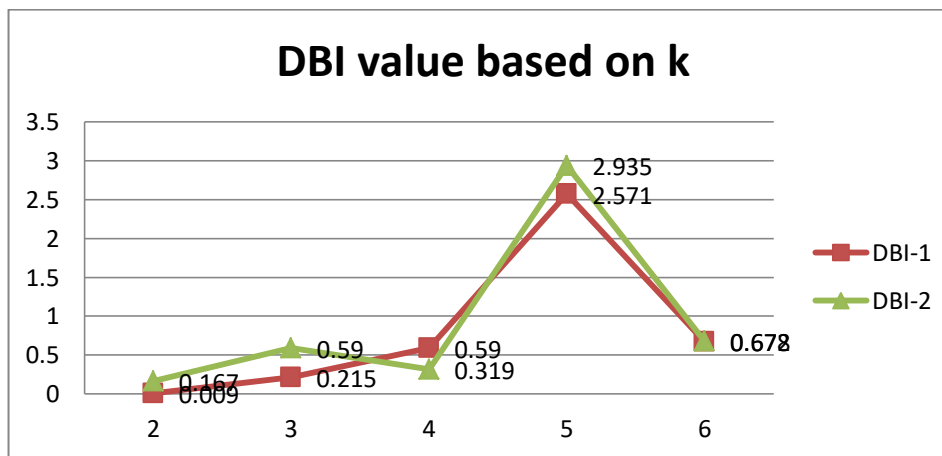| k	DBI-1	DBI-2
2	0.009	0.167
3	0.215	0.590
4	0.590	0.319
5	2.571	2.935
6	0.672	0.678

Figure 5. Graph of DBI values for each k

It is possible to deduce from the results that the number of clusters used is 3 (k=3), with a DBI value of 0.215. The determination is made by taking the average of the two DBI results (DBI-1 and DBI-2), with DBI-1 being 0.006 better than DBI-2. The number of k=2 was not chosen despite having the smallest value because the resulting cluster results were unstable, namely (cluster0:1 and cluster1:33). As a result, the second option, k=3, is chosen. The results of the k=3 cluster analysis using the k-medoids method with RapidMiner software are shown below.
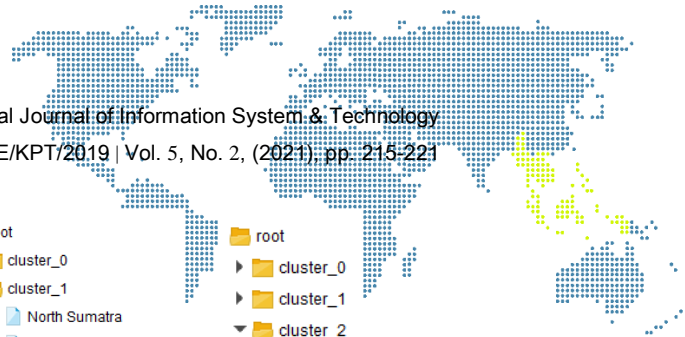
Cluster Model

```
Cluster 0: 20 items
Cluster 1: 13 items
Cluster 2: 1 items
Total number of items: 34
```

root
- ▼ cluster_0
  - Aceh
  - Riau
  - Jambi
  - South Sumatra
  - Bengkulu
  - Lampung
  - Kep. Bangka Belitung
  - West Kalimantan
  - Central Kalimantan
  - South Kalimantan
  - East Kalimantan
  - North Kalimantan
  - Central Sulawesi
  - Southeast Sulawesi
  - Gorontalo
  - West Sulawesi
  - Maluku
  - North Maluku
  - West Papua
  - Papua
- ▶ cluster_1
- ▶ cluster_2

root
- ▶ cluster_0
- ▼ cluster_1
  - North Sumatra
  - West Sumatra
  - Kep. Riau
  - DKI Jakarta
  - West Java
  - Central Java
  - DI Yogyakarta
  - East Java
  - Banten
  - West Nusa Tenggara
  - East Nusa Tenggara
  - North Sulawesi
  - South Sulawesi
- ▶ cluster_2

root
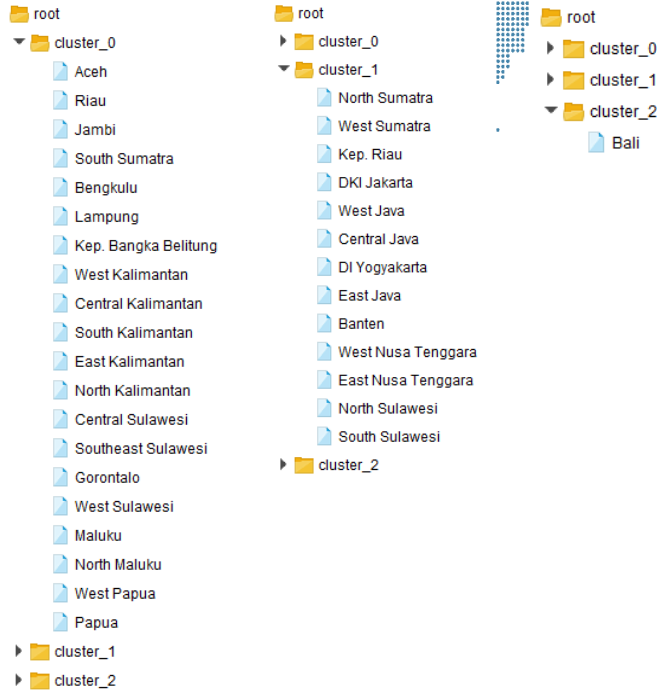- ▶ cluster_0
- ▶ cluster_1
- ▼ cluster_2
  - Bali

Figure 6. Results of cluster mapping with k-medoids.

Figure 6 depicts the clustering process based on the centroid value generated by the previous iteration. Cluster 0 is a low cluster with a centroid value of (11; 11; 13), cluster 1 is a normal cluster with a centroid value of (109; 78; 47), and cluster 2 is a high cluster with a centroid value of (5974; 9253; 5688). Figure 7 shows the final centroid result, and Figure 8 shows a visualization graph of the created cluster.

Attribute	cluster_0	cluster_1	cluster_2
2017	11	109	5974
2018	11	78	9253
2019	13	47	5688

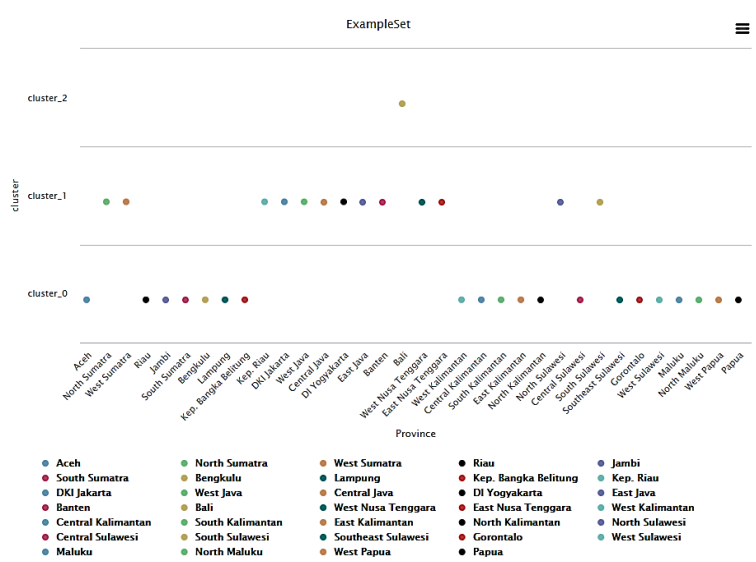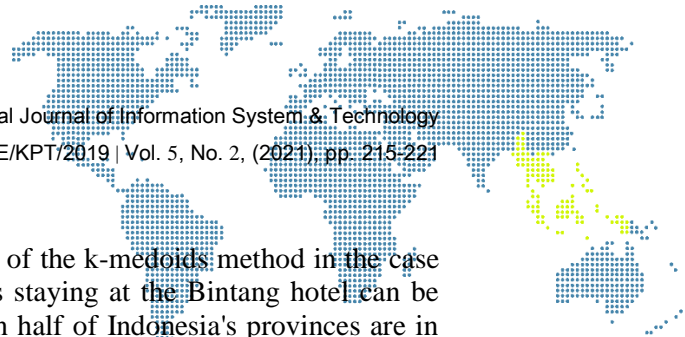**Figure 7.** Final centroid value

Figure 8. Cluster mapping graph

4. Conclusion

According to the study's findings, the application of the k-medoids method in the case of cluster mapping on the number of foreign guests staying at the Bintang hotel can be done with the formation of three clusters. More than half of Indonesia's provinces are in the low cluster, implying that an assessment is needed to ensure that foreign visitors visit more regions in Indonesia in order to increase Indonesia's tourism potential.

References

[1]   A. A. Rahma, "Potensi Sumber Daya Alam dalam Mengembangkan Sektor Pariwisata di Indonesia," *Jurnal Nasional Pariwisata*, vol. 12, no. 1, pp. 1–8, 2020.

[2]   M. Febriani Jayadi and I. B. Suryawan, "Strategi Pengembangan Potensi Pariwisata di Pantai Blimbingsari Kabupaten Banyuwangi," *Jurnal Destinasi Pariwisata*, vol. 8, no. 1, p. 10, 2020, doi: 10.24843/jdepar.2020.v08.i01.p02.

[3]   T. I. dan P. Direktorat Statistik Keuangan, "Tingkat Penghunian Kamar Hotel 2020," Badan Pusat Statistik, 2020.

[4]   Z. R. S. Elsi *et al.*, "Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia," *Journal of Physics: Conference Series*, vol. 1594, no. 1, 2020, doi: 10.1088/1742-6596/1594/1/012007.

[5]   M. Widyastuti, A. G. Fepdiani Simanjuntak, D. Hartama, A. P. Windarto, and A. Wanto, "Classification Model C.45 on Determining the Quality of Custumer Service in Bank BTN Pematangsiantar Branch," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019, doi: 10.1088/1742-6596/1255/1/012002.

[6]   F. Rahman, I. I. Ridho, M. Muflih, S. Pratama, M. R. Raharjo, and A. P. Windarto, "Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country," *IOP Conference Series: Materials Science and Engineering*, vol. 835, no. 1, 2020, doi: 10.1088/1757-899X/835/1/012058.

[7]   S. Sundari, Karmila, M. N. Fadli, D. Hartama, A. P. Windarto, and A. Wanto, "Decision Support System on Selection of Lecturer Research Grant Proposals using Preferences Selection Index," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–8, 2019, doi: 10.1088/1742-6596/1255/1/012006.

[8]   A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, and A. P. Windarto, "Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses," *IOP Conference Series: Materials Science and Engineering*, vol. 874, no. 1, pp. 1–9, 2020, doi: 10.1088/1757-899X/874/1/012024.

[9]   W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, "C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject," *Journal of Physics: Conference Series*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012005.

[10]   S. Dsouza, J. D. Dsouza, and T. Vanitha, "Analysis of Algorithms Data K-Means and K-Medoids," *International Journal of Latest Trends in Engineering and Technology*, pp. 370–373, 2017.

[11]   I. C. Dewi, B. Y. Gautama, and P. A. Mertasana, "Analysis of Clustering for Grouping of Productive Industry by K-Medoid Method," *International Journal of Engineering and Emerging Technology*, vol. 2, no. 1, p. 26, 2017, doi: 10.24843/ijeet.2017.v02.i01.p06.

[12]   A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means," *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, no. Iccmc, pp. 306–310, 2020, doi:

10.1109/ICCMC48092.2020.ICCMC-00057.

[13] J. Xiao, J. Lu, and X. Li, "Davies Bouldin Index based hierarchical initialization K-means," *Intelligent Data Analysis*, vol. 21, no. 6, pp. 1327–1338, 2017, doi: 10.3233/IDA-163129.

[14] N. Azis, Y. Ali, R. Subekti, P. Junianto, and L. Diner, "Mapping study using the unsupervised learning clustering approach," *IOP Conference Series: Materials Science and Engineering PAPER*, 2021, doi: 10.1088/1757-899X/1088/1/012005.