

Analysis of Foot Traffic due to COVID-19 using Mobile Network Big Data

A-Jeong Lee¹, Tae-Rang Park², and Jeong-Woo Jwa^{3*}

^{1,2,3} Jeju National University, Jeju, Korea

E-mail: lcr02@jejunu.ac.kr

Abstracts: The COVID-19 pandemic has hit the tourism industry with a drop in the number of domestic and foreign tourists and loss of revenue. We can prepare for future pandemics by analyzing the impact of COVID-19 on tourist numbers before and after the outbreak. In this paper, we analyze the change in the foot traffic of tourists and resident population before and after the outbreak of COVID-19 using mobile communication company call records (CDR) and WiFi access log data. Analysis of changes in the foot traffic due to COVID-19 is analyzed based on the number of confirmed cases, recoveries, and deaths by administrative dong, daily, and gender. The hierarchical cluster analysis of distribution population for administrative dong (classified as $k = 5$ clusters) was performed to classify areas with high changes in foot traffic due to COVID-19. We also predict the change in the foot traffic due to the epidemic using the Xgboost (Extreme Gradient Boosting) algorithm.

Keywords: Big Data Analytics, COVID-19, Mobile Communication Big Data, Foot Traffic Analysis, Hierarchical Cluster Analysis.

1. INTRODUCTION

As the use of smartphones expands due to the rapid development of information technology, all behaviors of people are converted into data, and a huge amount of data is being produced at a rapid pace. As a result, it became possible to collect objective and empirical data in real time (T. I. Kwon & C.H. Lee, 2017), and the social demand for processing various information into new information through big data analysis also increased (S. R. Kim, 2012). The importance of big data is also growing in the tourism field, and through big data analysis of social media and mobile communication networks of tourists, it is possible to understand the perception of activities during travel (S. Y. Lee, 2020). Since 2016, Jeju Island has been taking a strategic approach to forecasting demand for the tourism industry through data collection, analysis, and utilization of tourists' activities during travel (T.H. Um & N.H., 2019). In addition, Jeju proposed a plan to utilize ICT-based big data by implementing a real-time tourist destination congestion analysis service in preparation for Covid-19. Therefore, it is possible to design travel products for personalized travel planner services using big data analysis of mobile communication networks (H.C. Kang & J.W. Jwa, 2018; M. Švarcmajer, Ž. Hederić, & N. N. Cvetković, 2021). The movement of tourists worldwide has been greatly affected by the ongoing covid19 pandemic (S. Gössling, 2020). Evaluating changes in tourist travel due to COVID-19 is important for sustainable tourism and policy development for post-pandemic tourism recovery (W. Zheng, 2022; W. Benhsain & S. Salhi, 2022). Therefore, the need to study the change in the movement of tourists before and after the pandemic using telecommunication big data is emerging to cope with the pandemic, which has become a new policy direction for the tourism industry.

In this study, we analyze the change in the foot traffic of Jeju residents and tourists due to COVID-19 by using the call data record (CDR) of the mobile communication network and the WiFi access data. We analyze the impact of the tourism industry due to the decrease in the floating population by region due to COVID-19 through the stratum cluster analysis by administrative district on the change in the floating population. In addition, we use the Xgboost (Extreme Gradient Boosting) algorithm to predict changes in the floating population due to epidemics.

Different organizations now have to deal with complexity and unpredictability due to the environment's and their operations' quick and rapid development. To live, maintain, and advance in such a setting, companies have been forced to become more adaptable to unavoidable changes, as well as to deal with changes successfully and successfully in their external environments.

2. RELATED LITERATURE

We analyze the change in the number of residents and tourists in Jeju Island before and after the outbreak of COVID-19 to analyze the impact of the COVID-19 pandemic on the tourism industry. Analysis of the change in floating population due to COVID-19 provides regions vulnerable to the pandemic and causes.

2.1. Impact of COVID-19 on Tourist Movement

Mobility is an important part of tourism (D. Zheng, 2021), which includes the spatial movement of tourists (N. Shoval & M. Lsaacson, 2007), allowing us to understand what role tourism mobility plays in time and space (A. Hardy, 2020). A study on the impact of the COVID-19 pandemic on the tourism industry focused on and considered changes in domestic tourists (M. Ren, 2022). Y. Xu et al analyzed the impact of the pandemic on the tourism industry and concluded that when the risk of health disaster increases, tourism demand decreases (Y. Yang, 2020). In a study on the impact of the pandemic on tourism activities, a 10% increase in the pandemic index resulted in a 2.1% decrease in tourists (A. Forialdis, 2021; S. Polyzos, 2021). Unlike other activities, tourism shows considerable flexibility in tourism decision-making (Y.H. Hwang & D. R. Fesenmaier, 2011), and when the environment changes, different heuristics are made for various travel activities that differ from the perceived importance (S, Park, & D. Fesenmaier, 2014). Some recent tourism research results suggest that changes in traveler perceptions during a pandemic can affect travel behavior post-pandemic (Y.H. Hwang & D. R. Fesenmaier, 2011; W, Huang, 2020; Y. S. H. Dahami, 2022). In addition, infectious diseases have strengthened the tendency of travelers to avoid congestion in travel destinations and argued that it is important to avoid human contact with technical support such as self-service or AI service (J. Li, 2020; F. Iraguha, A. H. Ramelan, & P. Setyono, 2020; A. Akbar, 2022). This suggests that it is important to investigate behavioral changes in travel behavior after COVID-19 to identify changes in intimacy. In other words, by using mobility data, various activities of travelers can be investigated and changes in travel behavior can be analyzed in real time.

2.2. Impact of COVID-19 on the number of tourists in Jeju Island

Since the declaration of the COVID-19 pandemic, vaccines have been developed for COVID-19, but no cure has been developed, so the risk of an infectious disease remains. Recently, due to the influence of COVID-19, people's behavior patterns are undergoing major changes, and tourism activities are also affected (C. H. Park, 2020). In Jeju Island, the number of visitors decreased immediately after the outbreak of COVID-19, but as overseas travel was continuously restricted, the number of domestic travelers increased.

3. MOBILE COMMUNICATION BIG DATA AND ANALYSIS METHODS

We use mobile telecommunication big data before and after COVID-19 to analyze the change in Jeju Island's resident population and floating population of tourists. The floating population analysis uses the R big data analysis tool, and the Xgboost (Extreme Gradient Boosting) algorithm is used to predict the floating population.

3.1. Mobile Communication Big Data for Foot Traffic Analysis

In this paper, we analyze the foot traffic of Jeju Island tourists and Jeju residents before and after the outbreak of COVID-19 using call data record (CDR), which is data accessed by a smartphone to a base station in a mobile communication network, and WiFi access log data. Through this, we can understand the impact of the epidemic outbreak on the foot traffic and the impact of the epidemic on the tourism industry and the local economy. To analyze the foot traffic, we perform big data analysis on the data from January to April 2019, before the outbreak of COVID-19, and the four-month data from January to April 2020, immediately after the outbreak of COVID-19. The data used for big data analysis provides GRS80 coordinate data in 50mx50m cell (50cell) units, and population distribution of Jeju residents and tourists by day, time, gender, and age. Figure 1 shows the distribution of the number of tourists in 50 cell units on the GIS map. We can see the areas where tourists are concentrated as the red color on the map indicates the high tourist density. Figure 2 shows the process of analyzing the resident population and floating population of tourists using the R big data analysis tool. We use the legal code to preprocess the raw data in units of 50 cells into administrative units.

We use the Xgboost algorithm to predict changes in the floating population due to epidemics. Xgboost is a

distributed boosting ensemble machine learning algorithm optimized for advanced prediction performance (T. Chen & C. Guestrin, 2016). Xgboost is an ensemble learning algorithm generated using a boosting tree model that builds many layers and then aggregates them to form a single consensus prediction model (R. E. Schapire, 2003; L. Y. Muhammad, Husaini, E. Suhartono, E. Wydiamala, L. Marlinae, 2021). In general, these ensemble machine learning algorithms outperform non-ensemble algorithms such as logistic regression, vector machine support, and linear discriminant analysis (Ye. Pengfei, 2022).

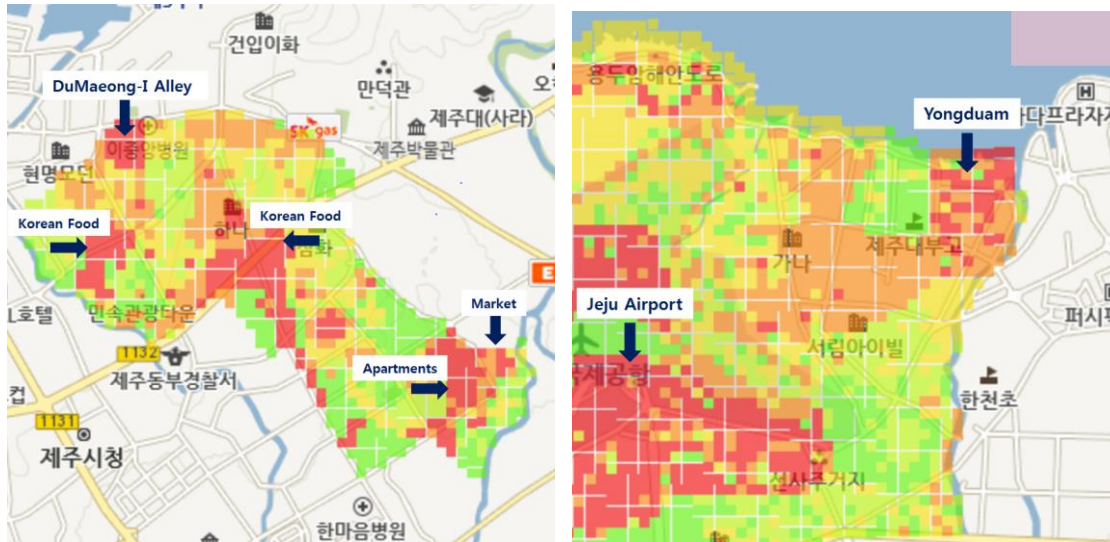


Figure 1. Distribution of the number of tourists in 50 cell (pixel) units on the GIS map.

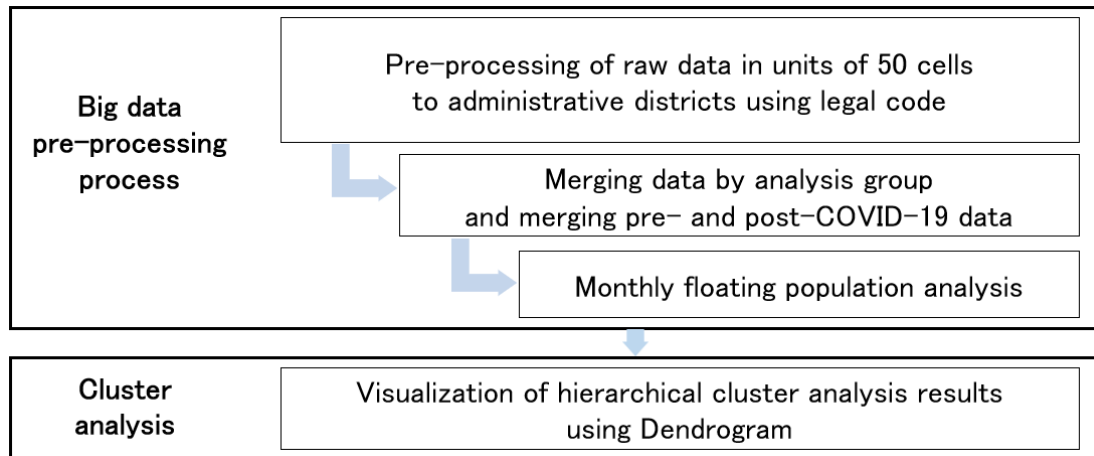


Figure 2. Foot traffic analysis procedure using R big data analysis tool.

3.2. Hierarchical Clustering Analysis for Foot Traffic Analysis

For the foot traffic analysis by administrative district, we perform preprocessing to combine population data in 50 cell units by administrative district. Analysis of the foot traffic of Jeju residents and tourists due to COVID-19 is performed by number of confirmed cases, time and month, gender, and age. To classify administrative dong with high changes in foot traffic due to COVID-19, a hierarchical clustering analysis (classified as $k=5$) was performed for administrative dong with similar changes in foot traffic. Through the hierarchical clustering analysis, it is possible to analyze the characteristics of regions where the foot traffic changes due to COVID-19. As a result of the foot traffic analysis, it is possible to confirm the difference in the amount of change in the foot traffic due to Corona 19 by region. Figure 3 shows the order of administrative districts where the decrease in floating population occurred immediately after the outbreak of COVID-19. In Chapter 4, hierarchical clustering analysis (classified as $k=5$) is used to analyze the area where the decrease in floating population occurs and the cause.

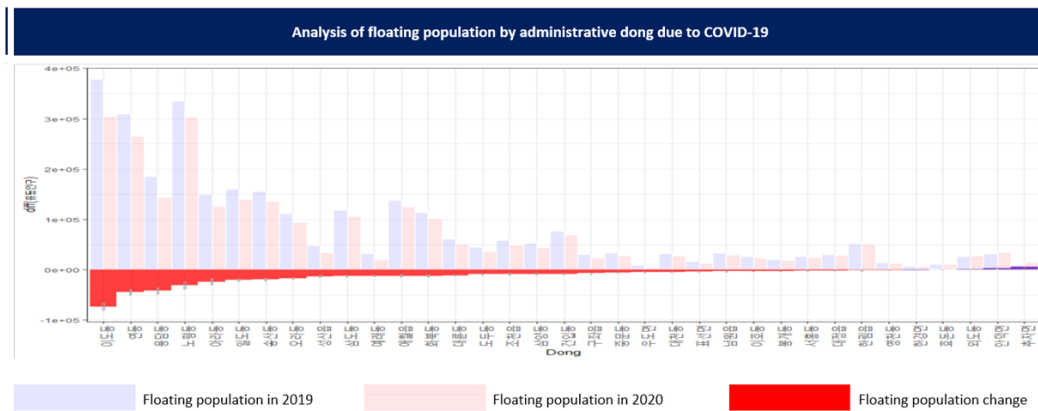


Figure 3. Floating population analysis results shown in order of administrative dong with high floating population decrease.

4. ANALYSIS RESULTS

We analyze the change in the floating population of tourists using big data of the mobile communication network for 4 months before and after the COVID-19 to analyze the impact of the COVID-19 outbreak on the tourism industry and the local economy. We perform a cluster analysis of administrative districts to identify regions that was significantly affected by COVID-19. In addition, the Xgboost algorithm is used to predict the change in the floating population due to epidemics. A clustergram was constructed to find out how the floating population and living population (inflow and outflow) activities changed during the period from January to April 2019 and from January to April 2020.

Figure 4 shows the cluster dendrogram for changes in floating population and the amount of change in floating population by administrative dong. As a result of the floating population analysis, after the outbreak of COVID-19, a decrease in floating population occurs in the order of Ido-dong, Yongdam-dong, Nohyeong-dong, and Yeon-dong, where the population density is high, and restaurants and accommodations are concentrated as shown in Fig 4. In the early stages of the COVID-19 pandemic, people avoid densely populated areas. Andeok-myeon, Namwon-eup, and Hangeong-myeon, where the population density is low and there are few accommodation facilities, rather increase the floating population. As a result of analyzing the inflow and outflow population before and after the outbreak of COVID-19, there is a relatively large outflow in densely populated areas and a large inflow in low-populated areas. This is a result of reflecting the health policy stance of social distancing since covid-19 is transmitted through droplets. In fact, within the same period, it can be seen in the same context as the decrease in tourism to places with high aggregation intensity such as urban tourist facilities, accommodations, and MICE, and the increase in nature/ecological tourist attractions and camping tourists. As a result of analyzing the floating population of Jeju, it can be seen that the floating population decreased in March and April from areas with high population density and many accommodation and restaurant facilities, and tourists moved to Myeon and Dong in administrative districts, which are nature and ecotourism destinations with low population density.

Figure 5 shows the change in floating population before and after the outbreak of the COVID-19 pandemic in terms of administrative dong, age, number of confirmed COVID-19 cases, and gender. As a result of the floating population analysis, the relatively active 20s and 30s show a significant decrease in the floating population due to COVID-19. As the number of COVID-19 confirmed cases increases in the early stages of the corona outbreak, the number of floating populations also decreases. The decrease in the floating population of women is relatively greater than that of men in comparison between men and women.

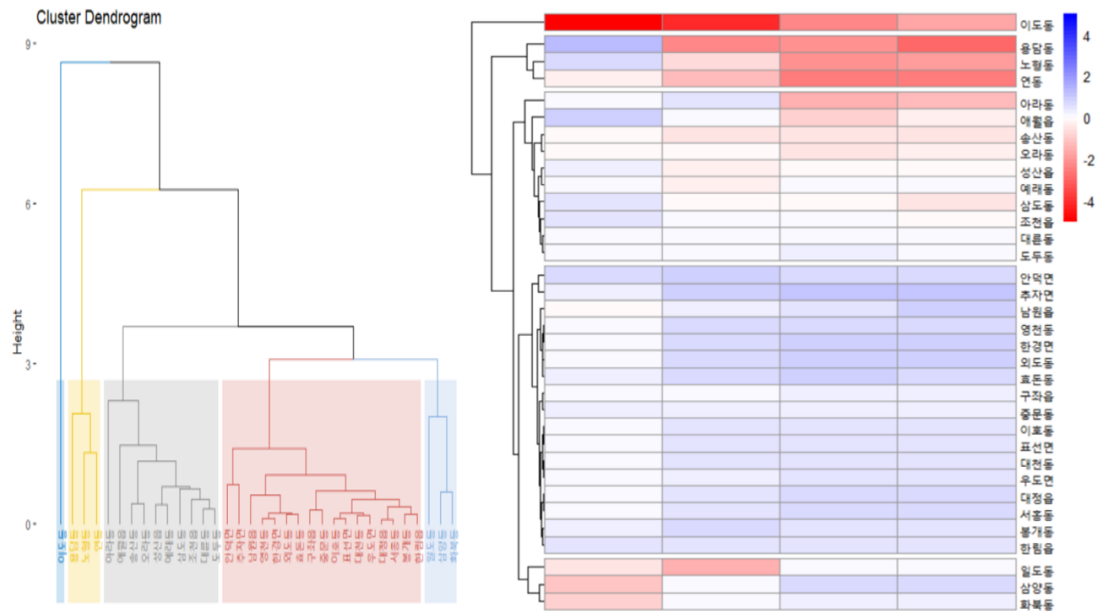


Figure 4. Cluster dendrogram and changes in floating population by administrative dong.

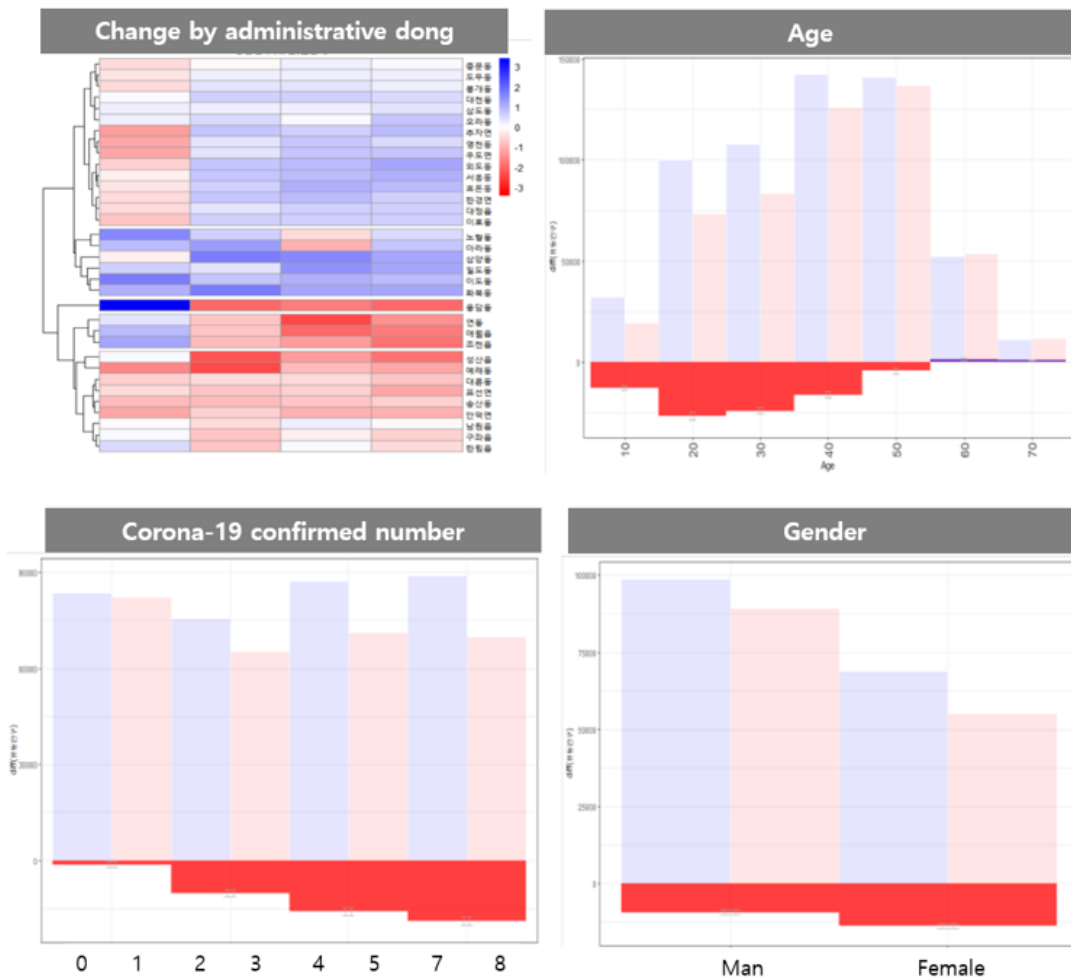


Figure 5. Changes in floating population before and after COVID-19.

We set the ratio of the Xgboost algorithm's train set and test set to 7:3 to predict the number of floating populations from big data. Figure 6 shows the floating population prediction results with the Bland-Altman plot and

the comparison between predicted and actual values. The data average, predicted average, and difference average of the floating population are $72,371 \pm 12,9178$, $72,208 \pm 128,106$, and $5,941 \pm 11,632$, respectively. The floating population prediction accuracy using the Xgboost algorithm is 91.79%. The accuracy of the prediction result for the inflow and outflow of the floating population is 95.19% and 94.26%, respectively.

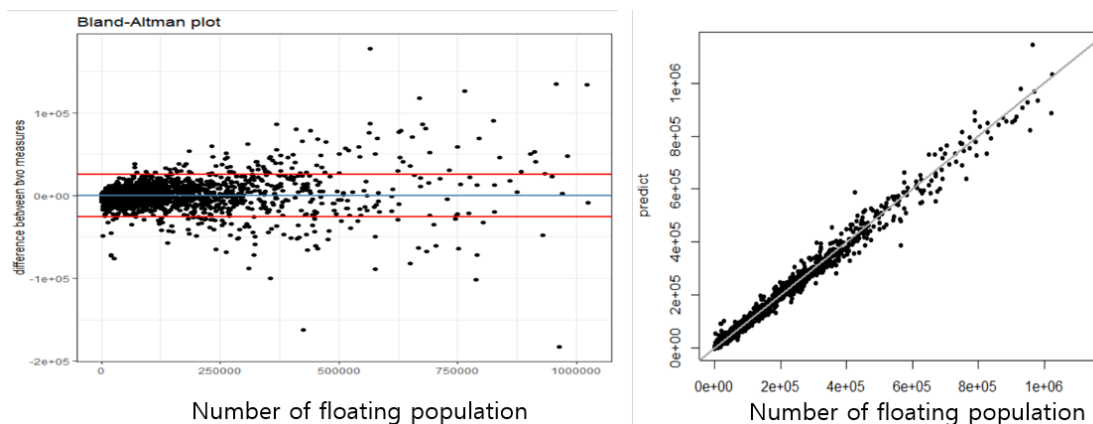


Figure 6. Bland-altman plot and comparison of prediction and actual data.

5. CONCLUSIONS AND DISCUSSIONS

The outbreak of the COVID-19 pandemic has drastically reduced the number of tourists to Jeju Island, Korea's representative tourist destination, and has dealt a huge blow to the tourism industry. In this paper, we analyze the change in the floating population of tourists using mobile communication big data of 4 months before and after the COVID-19 to analyze the impact of the COVID-19 outbreak on the tourism industry and the local economy. We analyze the floating population, inflow, and outflow population by administrative district in Jeju Island by the number of confirmed COVID-19 cases, gender, age, month, and hour before and after the outbreak of COVID-19. As a result of Hierarchical Clustering Analysis, there was a decrease in floating population and population outflow in areas with high population density and those in their 20s and 30s. The floating population and inflow population increase in areas with relatively low population density and many nature and ecotourism sites. We predict floating population, inflow and outflow population using the Xgboost algorithm. The prediction accuracy of the floating population, inflow, and outflow population of the Xgboost algorithm is 91.79%, 95.19%, and 94.26%, respectively. We plan to use credit card company big data to analyze the impact of the COVID-19 pandemic on the local economy and analyze tourist movement patterns using mobile communication big data.

6. ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2021R1A2C1093283).

REFERENCES

- [1] T. I. Kwon, & C.H. Lee, "A Study on the Utilization System and Empirical Analysis of Big Data in Tourism". *Policy Research: Korea Culture & Tourism Institution*, Vol. 2017.
- [2] S. R. Kim, M. M. Kang, & S. M. Park." The Future of Big Data". *Communications of the Korean Institute of Information\ Scientists and Engineers*, Vol.30, No.6, pp.18-24, 2012.
- [3] S. Y. Lee, T. K. Kim, & N. H. Chung, Consumers' Contribution Behaviors in Seoul Tourism Contents: Relationship between Like, Comment and Sharing in Chinese Social Media. *The Journal of Internet Electronic Commerce Research*. Vol.20, No.3, pp.17-31, 2020.
- [4] T.H. Um, & N.H. Chung, "Does smart tourism technology matter? Lessons from three smart tourism cities in South", *Asia Pacific Journal of Tourism*, VOL. 26, NO. 4, pp.396-414, 2019.
- [5] H.C. Kang, & J.W. Jwa, "Development of Android Based Smart Tourism Application Based on Tourism Bigdata Analytics", *Journal of Engineering and Applied Sciences*, Vol. 13, no. 5, pp. 1164-1169, 2018.
- [6] S. Gössling, D. Scott, & CM. Hall, "Pandemics, tourism and global change: A rapid assessment of COVID-19", *Journal of Sustainable Tourism*, Vol. 29, No. 1, pp.1-20, 2020.
- [7] W. Zheng, M. Li., Z. Lin, & Y. Zhang, "Leveraging tourist trajectory data for effective destination planning and management: A new heuristic approach", *Tourism Management*, Vol. 89, 2022.

- [8] D. Zheng, Q. Luo, & B. W. Ritchie, "Afraid to travel after COVID-19? Self-protection, coping and resilience against pandemic 'travel fear'", *Tourism Management*, Vol. 83, 2021.
- [9] N. Shoval, & M. Lsaacson, "Tracking tourists in the digital age", *Annals of Tourism Research*, Vol. 34, No. 1, pp141-159, 2007.
- [10] A. Hardy, A. Birenboim & M. Wells., "Using geoinformatics to assess tourist dispersal at the state level", *Annals of Tourism Research*, Vol 82, 2020.
- [11] Dahami, Y. S. H. (2022). Home in the Poetry of Saudi Arabia Poets: Mustafa Balilah an Example of a Distinguished Arab (1). *International Journal of Humanities, Arts and Social Sciences*, 8(1), 1-8.
- [12] M. Ren et al., "Impact of the COVID-19 pandemic on travel behavior: A case study of domestic inbound travelers in Jeju, Korea". *Tourism Management*, Vol 92. 2022.
- [13] Y. Yang, H. Zhang, & X. Chen., "Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak". *Annals of Tourism Research*, Vol. 83, 2020.
- [14] A. Forialdis, S. Polyzos, & T.C.T. Huan., "The good, the bad and the ugly on COVID-19 tourism recovery", *Annals of Tourism Research*, Vol. 87, 2021.
- [15] S. Polyzos, A. Samitea, & A. E. Spyridou., "Tourism demand and the COVID-19 pandemic: An LSTM approach", *Tourism Recreation Research*, Vol. 46, No. 2, pp175-187, 2021.
- [16] S, Park, & D. Fesenmaier, "Travel decision flexibility", *Tourism Analysis*, Vol. 19, No. 1, pp35-49. 2014.
- [17] Y.H. Hwang, & D. R. Fesenmaier, "Unplanned tourism attraction visits by travelers", *Tourism Geographies*, Vol. 13, No. 3, pp.398-416. 2011.
- [18] L. Y. Muhammad, Husaini, E. Suhartono, E. Wydiamala, L. Marlinae (2021). Meta analysis: Relationship of recapping actions and years of service with the incidence of needle stick injury in health workers at health facilities. *Journal of Advances in Health and Medical Sciences*, 7
- [19] Iraguha, F., Ramelan, A. H., & Setyono, P. (2020). The influence of environmental factors on COVID-19: Review. *Journal of Applied and Physical Sciences*, 6(1)
- [20] Benhsain, W., & Salhi, S. (2022). The territorial governance of the COVID-19 sanitary crisis and the challenges of the tourism actors in the old Medina of Marrakech. *Journal of Advances in Humanities and Social Sciences*, 8(2), 67-76.
- [21] W, Huang et al, "Twitter reveals human mobility dynamics during the COVID-19 pandemic", *PLoS One*, Vol. 15, No. 11, 2020.
- [22] J. Li, T. H. H. Nguyen, & J. A. Coca-Stefaniak, "Coronavirus impacts on post-pandemic planned travel behaviours". *Annals of Tourism Research*. 2020.
- [23] C. H. Park, "Untact Economy and Changes in Business Eco-System, Implication for Inclusive Growth Policy - Impact of e-commerce on job and regional balance", *The Journal of Internet Electronic Commerce Research*, VOL. 20, NO. 5, pp.181-200, 2020.
- [24] T. Chen, & C. Guestrin, "Xgboost: a scalable tree boosting system". *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 85–94, 2016.
- [25] A. Akbar (2022). Controlled and Autonomous Motivation Relation with Social Media Network Content Contribution: Post-COVID-19 Scenario. *Journal of Digitovation and information system*, 2(2), 55-74.
- [26] R. E. Schapire, "The boosting approach to machine learning: an overview", *Nonlinear Estimation and Classification*. pp. 149–171, 2003.
- [27] Ye. Pengfei, et al., "Diagnosis of obstructive sleep apnea in children based on the XGBoost algorithm using nocturnal heart rate and blood oxygen feature", *American Journal of Otolaryngology*, 2022.
- [28] Švarcmajer, M., Hederić, Ž., & Cvetković, N. N. (2021). High Voltage Power Distribution Network Insulator Fault Monitoring Using IoT Technologies. *International Journal of Technology and Engineering Studies*, 7(2).

DOI: <https://doi.org/10.15379/ijmst.v10i1.1453>

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.