

Using Machine Learning to Discover Tourism Areas of Interest from Tweet Clusters

B. Devkota¹ and H. Miyazaki²

¹Department of ICT, Asian Institute of Technology, ²Center for Spatial Information Science, University of Tokyo
Contact: <bidur.devkota@gmail.com, heromiya@csis.u-tokyo.ac.jp>

(1) **Introduction:** This study utilizes free data sources (like OSM, NTL and Twitter) and apply machine learning to detect tourist spots in tourism districts of Nepal. Four different machine learning classifiers were trained and their performance was examined.

(2) **Method:** Figure 1 shows the machine learning pipeline. Firstly, spatial tweet clusters were obtained using DBSACN algorithm (Devkota et. al., 2019) and sixteen different input features were collected from those clusters. Nine relevant features i.e. principal components are extracted by applying dimensionality reduction algorithms. Next, training data is fit using four classifiers (Support Vector Machine, Logistic Regression, Random Forest and Multilayer Perceptron Neural Network). The k-fold cross-validation was followed in order to minimize bias in the final model. Finally, the model performance was examined using the test dataset.

(3) **Result:** Figure 2 shows the performance of the selected classifiers based on different number of principal components. The best results were obtained while using 8 principal components, i.e. F1 score of 0.82 was obtained with SVM and Random Forest algorithms.

The results of this study illustrated that the proposed approach seems promising in identifying major tourist attractions the Kathmandu valley. SVM is preferable for obtaining quick results in data/resource constrained scenarios. However, RF proved more preferable as it does not require preprocessing steps and can adapt to new data with least or no performance degradation.

(4) **Data:**

- Twitter Data (2017-2019)
- OpenStreetMap Data for Nepal (2019)
- WorldPop (2018)
- VIIRS NTL Annual Composite Data (2016)

(5) **Acknowledgments:** This study was supported by the Japanese Government Scholarship at Asian Institute of Technology, Thailand.

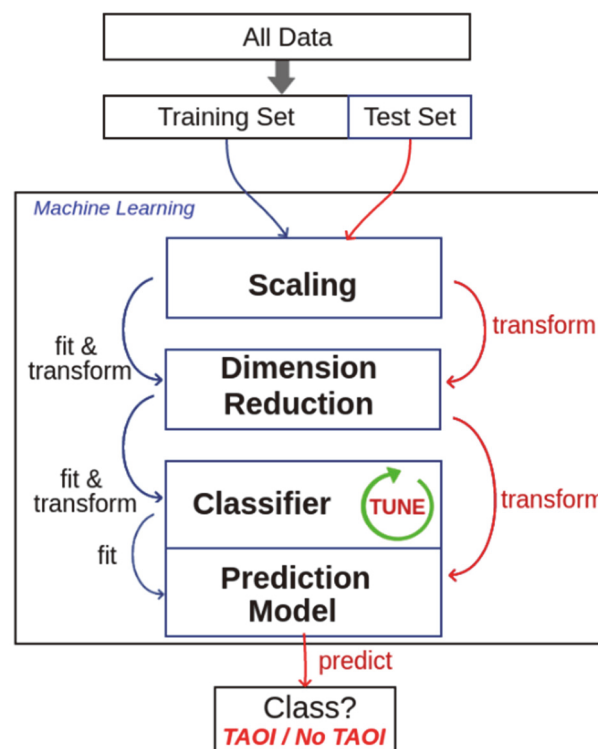


Figure 1: Approach to automate tourist spot discovery



Figure 2: F1 Score for different machine learning classifiers

(6) **Reference:**

Devkota, B., Miyazaki, H., Witayangkurn, A., & Kim, S. M. (2019). Using Volunteered Geographic Information and Nighttime Light Remote Sensing Data to Identify Tourism Areas of Interest. *Sustainability*, 11(17), 4718