

## Homeless Population Modeling and Trend Prediction Through Identifying Key Factors and Machine Learning

**Authors :** Shayla He

**Abstract :** Background and Purpose: According to Chamie (2017), it's estimated that no less than 150 million people, or about 2 percent of the world's population, are homeless. The homeless population in the United States has grown rapidly in the past four decades. In New York City, the sheltered homeless population has increased from 12,830 in 1983 to 62,679 in 2020. Knowing the trend on the homeless population is crucial at helping the states and the cities make affordable housing plans, and other community service plans ahead of time to better prepare for the situation. This study utilized the data from New York City, examined the key factors associated with the homelessness, and developed systematic modeling to predict homeless populations of the future. Using the best model developed, named HP-RNN, an analysis on the homeless population change during the months of 2020 and 2021, which were impacted by the COVID-19 pandemic, was conducted. Moreover, HP-RNN was tested on the data from Seattle. Methods: The methodology involves four phases in developing robust prediction methods. Phase 1 gathered and analyzed raw data of homeless population and demographic conditions from five urban centers. Phase 2 identified the key factors that contribute to the rate of homelessness. In Phase 3, three models were built using Linear Regression, Random Forest, and Recurrent Neural Network (RNN), respectively, to predict the future trend of society's homeless population. Each model was trained and tuned based on the dataset from New York City for its accuracy measured by Mean Squared Error (MSE). In Phase 4, the final phase, the best model from Phase 3 was evaluated using the data from Seattle that was not part of the model training and tuning process in Phase 3. Results: Compared to the Linear Regression based model used by HUD et al (2019), HP-RNN significantly improved the prediction metrics of Coefficient of Determination ( $R^2$ ) from -11.73 to 0.88 and MSE by 99%. HP-RNN was then validated on the data from Seattle, WA, which showed a peak %error of 14.5% between the actual and the predicted count. Finally, the modeling results were collected to predict the trend during the COVID-19 pandemic. It shows a good correlation between the actual and the predicted homeless population, with the peak %error less than 8.6%. Conclusions and Implications: This work is the first work to apply RNN to model the time series of the homeless related data. The Model shows a close correlation between the actual and the predicted homeless population. There are two major implications of this result. First, the model can be used to predict the homeless population for the next several years, and the prediction can help the states and the cities plan ahead on affordable housing allocation and other community service to better prepare for the future. Moreover, this prediction can serve as a reference to policy makers and legislators as they seek to make changes that may impact the factors closely associated with the future homeless population trend.

**Keywords :** homeless, prediction, model, RNN

**Conference Title :** ICSW 2022 : International Conference on Social Work

**Conference Location :** Tokyo, Japan

**Conference Dates :** September 08-09, 2022