A Multidimensional Study on the Determinants of Homelessness in Urban America

Jasmine Jia

Abstract

Homelessness remains a persistent issue in the United States, driven by economic disparities, housing crises, and systemic inequities. This study employs machine learning models and statistical analysis to investigate the primary factors influencing homelessness rates in urban areas. Using Point-in-Time (PIT) Count and Housing Inventory Count (HIC) data, alongside socio-economic indicators from the U.S. Census Bureau and the Bureau of Labor Statistics, this research examines the impact of income levels, unemployment rates, housing availability, and racial demographics. By leveraging advanced regression models and tree-based machine learning techniques, this study aims to improve predictive accuracy and inform policy interventions targeting homelessness prevention.

Introduction

Homelessness continues to challenge urban centers across the United States, with over 580,000 individuals reported as homeless in 2023. The crisis is exacerbated by rising housing costs, economic downturns, and systemic barriers that disproportionately affect marginalized populations. Policy efforts require a data-driven understanding of the factors that contribute to homelessness in order to implement effective interventions. This study investigates the relationship between homelessness and socio-economic indicators, seeking to understand the predictive power of different variables through machine learning techniques. The research employs multiple statistical approaches, including traditional regression models and advanced ensemble learning methods, to capture non-linear relationships and interactions between predictors. The goal is to generate insights that can guide targeted policy measures for reducing homelessness.

Data and Methodology

This research relies on two primary datasets: the Point-in-Time (PIT) Count, which provides an annual snapshot of homelessness populations, and the Housing Inventory Count (HIC), which details available housing resources such as emergency shelters and permanent supportive housing. Additional socio-economic variables, including median household income, unemployment rates, median home values, and racial composition, were obtained from the U.S. Census Bureau and the Bureau of Labor Statistics. A key challenge in homelessness research is the presence of high-dimensional, complex relationships between explanatory variables and homelessness rates. To address this, the study applies a combination of traditional and machine learning methods. The baseline model is a linear regression, used to establish fundamental associations between predictors and homelessness rates. However, given the limitations of linear models in capturing complex, non-linear interactions, advanced tree-based methods such as Random Forest and Gradient Boosting Machines (GBM) were incorporated. These models offer improved predictive performance by allowing for flexible interactions between independent variables and reducing bias-variance trade-offs. To validate the robustness of the models, cross-validation techniques were employed, ensuring that findings were not sensitive to overfitting. Model performance was evaluated using R^2 scores and Mean Squared Error (MSE). The feature importance scores derived from tree-based models provided insights into the relative influence of different socio-economic factors on homelessness rates.

Results and Analysis

The analysis identified key predictors of homelessness across urban areas. The results indicate that lower median household income significantly correlates with higher homelessness rates, suggesting that economic vulnerability is a major driver of homelessness. Additionally, areas with higher unemployment rates exhibited increased homelessness prevalence, reinforcing the importance of labor market stability in preventing housing displacement. Housing affordability emerged as a crucial determinant, with median home values showing a positive correlation with homelessness rates. This finding aligns with existing research that links high housing costs to increased housing insecurity. The number of available supportive housing units demonstrated a mitigating effect, with regions offering greater permanent supportive housing resources experiencing lower homelessness rates.

Racial disparities in homelessness were also evident, with Black populations disproportionately affected. This underscores the systemic inequities that contribute to homelessness among minority communities, emphasizing the need for targeted policy measures to address racial disparities in housing access. In terms of model performance, Random Forest and Gradient Boosting Machines (GBM) outperformed linear regression models, achieving R^2 scores above 0.98. These machine learning models successfully captured complex interactions between variables, revealing non-linear relationships that traditional models failed to detect. The inclusion of feature importance scores allowed for a more interpretable analysis, highlighting the most influential variables in predicting homelessness rates.

Limitations

Despite its contributions, this study has several limitations. First, racial representation in the dataset was limited primarily to Black and White populations, excluding other racial and ethnic groups that may experience unique housing challenges. Additionally, the accuracy of PIT Counts remains a concern, as these surveys provide a single-night estimate that may undercount transient homeless populations.

Machine learning models, while powerful, present their own challenges. Tree-based models are prone to overfitting, requiring extensive hyperparameter tuning and cross-validation to ensure robustness. Moreover, the dataset lacked certain social service and mental health indicators, which could further enhance predictive accuracy. Future research should incorporate these factors to create a more holistic understanding of homelessness determinants.

Policy Implications and Recommendations

The findings of this study underscore the need for comprehensive policy interventions to address homelessness. Expanding housing resources is a critical priority, with increased funding needed for permanent supportive housing and Housing First initiatives. These approaches have been shown to reduce homelessness by prioritizing stable housing before addressing other social or economic challenges. Economic policies should focus on enhancing labor market opportunities and providing job training programs for at-risk populations. Rental assistance programs for low-income individuals can help prevent homelessness by bridging the gap between wages and rising housing costs.

Addressing racial disparities in homelessness requires further research into systemic inequities that contribute to housing insecurity among minority communities. Equitable urban policies should be developed to prevent housing discrimination and expand access to affordable housing for historically marginalized groups.

Finally, improvements in data collection and analysis are necessary for effective policy implementation. Real-time homelessness tracking systems should be integrated with machine learning insights to provide more accurate and timely data for policymakers. This would allow for adaptive interventions that respond dynamically to changes in homelessness trends.

Conclusion

This study provides a data-driven examination of the determinants of homelessness in urban America, leveraging machine learning techniques to uncover key socio-economic and housing-related predictors. The results highlight the significance of economic stability, housing affordability, and racial disparities in shaping homelessness rates. The superior performance of tree-based models demonstrates the value of machine learning in policy analysis, offering deeper insights into complex social issues.

Future research should expand on these findings by incorporating additional demographic factors and exploring longitudinal data to capture trends over time. By integrating advanced analytical methods with targeted policy interventions, stakeholders can develop more effective strategies to combat homelessness and promote housing stability in urban communities.