Predictive Modeling for Uninsured Rates in the United States

: A Comprehensive Analysis

Sri Sai Durga Katreddi  
11626267

Sri Chandra Dronavalli  
11585371

Sireesha Bollu  
11561931

Information Science  
28th November 2023

**PROBLEM STATEMENT**

In the complex landscape of healthcare accessibility and affordability, predicting the uninsured rate has emerged as a pivotal task. The uninsured rate serves as a compass guiding decision-makers in resource allocation, policy formulation, and the implementation of targeted interventions. Our objective is to craft a robust predictive model that accurately forecasts the uninsured rate in the United States. This endeavor is not merely an analytical exercise; it is a crucial step towards influencing positive change in healthcare policy, reducing the number of uninsured individuals, and fostering equitable healthcare access for all.

**APPROACH**

Our journey begins with an exhaustive exploration of a dataset enriched with variables spanning years, FIPS codes, county and state information, and various demographic factors. The data preprocessing phase is meticulous, addressing missing values, converting categorical variables, and engineering new features, such as segmented uninsured rates for different age groups. Subsequently, we meticulously partition the dataset into input features (X) and the target variable (y), setting the stage for the development of predictive models.

An array of regression algorithms, including Decision Tree Regressor, Linear Regression, Random Forest Regressor, K-Neighbors, Bagging Regressor, Linear SVR, Lasso, AdaBoost Regressor, and CatBoost Regressor, are brought into play. Among these, the Random Forest Regressor stands out, exhibiting commendable performance in discerning underlying patterns and relationships, especially in the face of outliers and noise.

Our analysis extends beyond algorithmic intricacies, delving into the realms of exploratory data visualization. Through visual insights, we unravel the distribution of uninsured rates across diverse states, metropolitan and non-metropolitan areas, and distinct age groups. Correlation matrices and statistical analyses further enrich our understanding of the intricate relationships between variables, providing a holistic perspective on the factors influencing the uninsured rate.

**CONCLUSIONS**

As we unveil the curtain on our analysis, the Random Forest Regressor emerges as the torchbearer, illuminating the path to accurate uninsured rate predictions. The exploration not only sheds light on average uninsured rates per state but also unveils the differential impact of metro/non-metro classifications and traces trends over time for various age groups. Noteworthy is the identification of states with maximum and minimum uninsured rates, laying the groundwork for targeted interventions.

In the realm of model comparison, the emphasis is on selecting an algorithm that strikes the delicate balance between accuracy and robustness. The Random Forest Regressor triumphs in this context, positioning itself as the vanguard of predictive modeling for uninsured rates.

The implications of our analysis extend far beyond the confines of data. Policymakers, healthcare providers, and insurers can leverage these results to make informed, data-driven decisions. The insights garnered from our comprehensive analysis are poised to steer strategies that address healthcare disparities and uplift vulnerable populations, heralding a positive transformation in healthcare access across the nation.

**Team Members and Roles**