

Healthcare infrastructures in the United States have been debated for long time and were recently reformed to provide more affordable service to the public. The government noted that one of the drivers of high healthcare premiums was high readmission rate in some hospitals. Thereby, the reform implemented the readmission score to evaluate hospital performances. A high readmission rate not only indicates that the initial treatment is inappropriate or insufficient, but also forces insurance companies to compensate for unexpected expenses. Consequently, the cost of healthcare premiums inevitably increases, making healthcare more expensive and unaffordable.

To investigate the readmission rate issue, I hypothesize that there may be a relationship among several regional financial factors (such as personal income, poverty, government revenue and expenditures) and the readmission rate. In other words, low finance may correlate with low hospital performance because such hospital may not be able to supply or equip sufficient materials/stuff to serve patients. An outcome of interest is hospital performance per county (outcome = *better*, *national average*, or *worse*) evaluated by the aggregated score of within-30-day readmission and death rates in four categories: heart attack, heart failure, pneumonia, and overall-readmission rate per hospital.

I encountered a problem of imbalanced numbers of outcomes: approximately 75% of the outcome was *national average*, 16% *better*, and 9% *worse*. To mitigate this, I trained my models with downsampled *national average* data, while the others were unmodified.

A simple model based on financial factors gave us some insights into ties between expected hospital performance and regional financial status. The final model was able to predict hospital performance per county with an average accuracy of 80% (three-class classification), whereas the classification of two extremes (*better* vs. *worse*) can be predicted with an accuracy of 90%. This suggests that financial status of the extremes differ noticeably. The results from three-class classification suggest that non-financial factors may contribute to hospital performance within the same financial range. Additionally, close examination of incorrectly classified data from the two-class classification can further inform valuable non-financial factors in hospital performance.

As a conclusion, predictions by the current model are not yet sufficient. For example, downsampling can induce distortion of modeling and results due to probable inaccurate representation of *national average* data. Thus, to improve modeling and/or represent data with more authenticity, for example, ensemble prediction may be adequate, where predictions are made by several models and a final prediction is determined by averaging those predictions.