**Predicting Impatient length of stay in hospitals for New York Area**

**--based on CatBoost ML algorithm**

**Naixin Zhang**

**Introduction**

**Objective**

The prediction of length of stay for patients in hospitals has been always considered a challenging task since it plays a key role in the operational success of hospitals in general. Building models that assist in prediction of patients’ LOS from their admission to their discharge is an extremely handy tool that health care providers can utilize.

The main purpose of this project is to predict the length of stay for the population in New York Area. Based on the characteristics of the given dataset, there are 37 features in the dataset, and 3 of them are continuous numerical variables and 11 of them are category variables. For the dataset with many categorical features the accuracy of CatBoost would be better compared to other ML algorithms. At the same time, we will do the feature selection to identify which are the significant features that are critical for length of stay prediction.

**Data description**

**data pre-processing**

**Methodology**

**Feature selection**

**classification method**

**Result**

**Conclusion and Future work**