PREDICTING THE COMMON VIOLATIONS OF CHILDCARE CENTERS AND SUMMER CAMPS IN NEW YORK

Problem Statement

New York State has regulations that govern the minimum requirements for licensed and registered childcare centers in the state of New York. These requirements have been developed to provide a healthy and safe environment for the children. These regulations are also designed to monitor the curriculums of the childcare centers. However, according to the National Institutes of Health, just 10% of child-care centers provide high-quality care[1]. There are 1,789,069 children under 18 years old in New York City. Raising a child, in such an expensive city necessitates both parents working full time. One of the biggest concern of a working parent is to provide high-quality childcare for her/his dependents. Typically, parents, who seek a reliable child care center, utilize two means: (1) parents look for reliable online reviews; (2) they would like to access the most recent inspection report from the state. In this project, I will investigate violations and regulations of child care centers and summer camps. The predictive model will be able to predict common health violations and observed problems of facilities.

Dataset

The data I will use is the DOHMH Childcare Center Inspections data that is provided by New York City. (https://catalog.data.gov/dataset/dohmh-childcare-center-inspections) This dataset contains a list of all inspections conducted and any associated violations at active, city-regulated, center-based child care programs and summer camps over the past three years. The data covers a period from 2016 up through 2019. It is provided as a CSV file, an RDF file, a JSON file, and an XML file.

Scope of the Project

The goal of the project is to develop a predictive model which can determine and identify the high-risk childcare centers based on the health code violations data in the state of New York. Possible stakeholders for this project are: Parents, childcare services, private and public schools, community centers, religious institutions are this project's potential stakeholders.

Technical Approach

- **Method:** Health code violations are measurements of the risk, which a childcare center constitutes to public safety. These measurements are based on observed deficiencies during the annual health inspections. Health code violations can include the misused sanitation or having employees without proof of immunization records. An inspector assigns a health code rating after reviewing

[1] https://www.nytimes.com/2019/05/24/opinion/child-care-crisis.html

all the hazards that are determined to be at critical control points. This study will analyze the violation rate percentages, violation status, and regulation summaries.

- **Working with Data:**
  After cleaning and wrangling the data, the project will focus on two violation rates(public health hazard and critical violations), and violation percentages for each facility. Keywords from inspection summaries will be extracted to categorize risk factors in four groups: employees, safety issues, health issues, and physical environment. The project will be capitalized on the EDA techniques to analyze the data. To be able to build a predictive model, a regression model will be implemented along with the machine learning techniques.
- **Model Creation:**

Project Outcomes

Project outcomes will include a jupyter notebook containing code, a report, and the slide deck.