Nosson Weissman
Professor James Topor
AIM 5001
March 26, 2021

AIM 5001 Final Project Proposal - Revised

Intro - Research Question

Creating an autonomous nurse-patient assignment system would be a unique way to demonstrate the potential of proper data management and machine learning. A quick google search yields research papers and even a long list of Git repos claiming to solve the nursing patient assignment problem. I will not be solving the nurse-patient assignment problem… But, on weekends I work at a group home for people with special needs. At the home, staff are assigned residents to work with and other tasks. These assignments are based on certifications completed as well as other specifics like years of experience, hours worked etc. Each shift, supervisors spend time figuring out a schedule (At the home, the paper with the staff-resident pairings is referred to as "the schedule" as it gives an outline for the schedule of the day) in which no one staff member is assigned a shift which is too difficult. While staff-resident assignments should involve questions similar to the nursing-patient assignment problem, the variables involved are generally different, and research is harder to come by. The long-term goal with this project is to modify the algorithms which exist for nurse-patient assignments to solve the staff-resident assignment problem. To begin, the current question, and one which even the nurse patient algorithms each answer differently, is **"What makes a schedule a 'good' or 'fair' schedule?"**.

Data to be Used

To answer this question, we will be using data both from staff scheduling info and past staff-resident assignments.

Staff Scheduling:

> We will be gathering staff scheduling data from a web app called WhenToWork, used by the home to track employee schedules. Data will be scraped using the Python Scraper Module   which has already been implemented thanks to Github user [mneps](). This step shall fulfill the requirement the *"Project includes at least one Python feature that we did not cover in class".* requirement.

Past Staff-Resident Assignments:

> Schedule data will be sourced directly from the group home in the form of written schedules. These paper (written and/or typed) schedules will be entered to Postgres database either manually or using the Python Tesseract library.

Approach

Schedule data will be loaded from the Postgres database into Pandas dataframes. WhenToWork data will be entered to Pandas dataframes and hours-worked will be used as a metric for employee aptitude/experience.

In creating the schedule, many variables must be accounted for. Some of the residents require specific certifications to work with them and specific shift duties can only be assigned to people with proper training and certifications (administering meds, driving company van, etc.), but also will not be assigned to a staff who is working with specific residents. The data values for each

part of the scheduling system will be broken down and analyzed, checking not only measures of central tendency (mean, median, mode etc.) but correlation between variables. We will use this data to deduce feature importance and the most important factors in creating a fair schedule. Visuals will be done using Mathplotlib.Pyplot.

This project is quite complex and has many parts. I have been working on this project on my own and I currently have already completed:

1. Schedule Data database (more comprehensive than the one seen in a previous project submission)
2. API for scraping whentowork (using the Flask framework and [mneps](#)' whentowork scraping code).
3. Python code for schedule data inserts to add new scheduling data to database
4. Have tested Tesseract on the paper schedules. And results look promising.
5. A clearly written outline of what needs to be done in preparation for this project (mainly having to do with data acquisition, but other things as well)

As such, although this would be complex to start from scratch, with the foundation in place this project is quite reasonable as a personal project.

Summary

The goal within the scope of "The Final Project" is to analyze existing schedule data and employee scheduling data to answer the question of what makes a fair schedule. Data will be collected from both the home and the employee scheduling system and analyzed using various Python libraries. We will focus on correlation of schedule attributes and feature importance. With a large part of the project foundation already built, the future of this project looks promising.