DTSC 691 Project Prooposal
A Machine Learning Approach to Opioid Prescription Crisis

Project Background

      The opioid crisis in the United States (U.S.) has grown significantly in size, and it is one of the most significant health care crises in recent years (Hodder et al., 2021). According to the U.S. Department of Health and Human Services (HHS), an estimated 10.1 million people misused opioid prescriptions in 2016. The opioid crisis in the U.S. can be traced back to the mid-1990s, when pharmaceutical companies assured the health care community that prescribing opioids for pain management has no impact on opioid addiction (Hancocks, 2019). As a result of this assurance by the pharmaceutical industry, there was an increase in opioid prescriptions, and the retail stock of opioids with retail pharmacies increased drastically from 105 million in 1998 to 210 million in 2010 (Hodder et al., 2021). According to an article by my addiction info (2019), an estimated sixty-four thousand people die of opioid drug overdose with 35% of all opioid deaths attributed to opioid prescription (Scholl et al., 2018).

      The resultant increment in opioid prescription created a crisis of addiction, and as a result, it is important that we examine the available dataset to see which healthcare practitioner will likely prescribe opioids the most. Answers to this question will aid government and interested stakeholders in developing policies that can help mitigate the opioid prescription crisis.

Goals of the Project

      This project aims to build a machine learning model capable of predicting the likelihood of a doctor prescribing opioids. The following objectives would be achieved:

1. understand the opioid dataset using Exploratory Data Analysis (EDA)
2. Build different classification models
3. Compare the model performance using different evaluation metrics
4. Select the best model based on the evaluation metrics performance

      The response variable for this research is the opioid prescriber, and the predictor variables are the various medications prescribed by the healthcare

practitioners. The research question I have developed for this project is: what is the probability of a medical practitioner prescribing opioids?

Data Description

The dataset for this project was retrieved from the archives of the UC Irvine machine learning repository (https://archive.ics.uci.edu/ml/datasets.php), which is a publicly available dataset and repository. The dataset contains medical practitioners' information: gender, specialization, list of prescriptions containing the different types of medication prescribed for other patients, and different cases. Specifically, the information about the different areas of medical practice/specializations is Urology, Surgeon, Dentist, Psychiatry, Nurse Practitioner, etc. The dataset contains information about the different types of prescription medication administered to patients, and the number of times it was prescribed.

In this project, the response variable is the opioid prescriber and has two outcomes (Yes, No). The predictor variables are the 239 medications prescribed by medical practitioners in their area of specialization. Also, the response variable is the opioid prescriber with two outcomes (Yes, No). The predictor variables are the 239 medications prescribed by a medical practitioner in the different healthcare specializations. The relationship between the response variable and the predictor variables is assumed to be a cause-and-effect relationship. This is so because prescribing a high dosage of opioid medication would likely result in addiction to the opioid medication.

Software
To complete my project, I will be utilizing python in Jupyter notebook, and several other packages complete my capstone project.

Analysis Plan
Analysis Description: Please see below for the analysis description.

Week 1 goals: My focus in week 1 is to complete my project proposal fully and then submit it to my mentor for review and feedback. Also, I will try to collect my dataset ready for week 2.

Week 2 goals: During week 2, I hope to have received some feedback from my mentor. I will incorporate any necessary changes and revise my proposal as required to ensure my proposal is ready for submission.

Week 3 goals: My goal for week 3 is to start working on my project. I will begin by importing the various libraries required for conducting Exploratory Data Analysis (EDA). Next, I will import my dataset, clean the data by filling missing values in features, and complete the Exploratory Data Analysis. I will also conduct some visualization and graphical representations of my dataset to gain insights and learn more about the variables in my dataset.

Week 4 goals: My focus in week 4 is encoding and dimensionality reduction. Since my dataset contains some categorical variables, it is necessary to transform categorical features into numeric features. Next, I will do feature importance and selection by removing less important features to address the issue of correlation issues.

Week 5 goals: During week 5, I will perform the model selection by building several models and then decide on a model that can generalize the training dataset and make accurate predictions.

Week 6 goals: After selecting my desired model, I will do hyperparameter tuning and interpret my model.

Week 7 goals: Complete a final review of my project, prepare a video presentation and submit my project for grading.
Delivery Plan: The delivery and presentation of my project will be completed using Jupyter Notebook. Utilizing Jupyter Notebook allows for a more engaging and interactive presentation.

References

Hancocks, S. (2019). The opioid crisis in the USA. *British Dental Journal*, *226*(11), 815-815. https://doi.org/10.1038/s41415-019-0420-6

Health and Human Services. (2017, December 4). *What is the U.S. opioid epidemic?* HHS.gov. https://www.hhs.gov/opioids/about-the-epidemic/index.html

Hodder, S. L., Feinberg, J., Strathdee, S. A., Shoptaw, S., Altice, F. L., Ortenzio, L., & Beyrer, C. (2021). The opioid crisis and HIV in the USA: Deadly synergies. *The Lancet*, *397*(10279), 1139-1150. https://doi.org/10.1016/s0140-6736(21)00391-3

My addiction info. (2019). *Opioid crisis and opioid epidemic in the U.S. -myaddictioninfo*. My Addiction Info. https://myaddictioninfo.com/the-opioid-crisis/#Key_Facts_about_the_Opioid_Crisis

Scholl, L., Seth, P., Kariisa, M., Wilson, N., & Baldwin, G. (2018). Drug and opioid-involved overdose deaths — United States, 2013–2017. *MMWR. Morbidity and Mortality Weekly Report*, *67*(5152). https://doi.org/10.15585/mmwr.mm675152e1