Raj Patel

DSI

07/16/18

**Capstone Technical Write Up**

For my capstone project, I’ve decided to tackle the major issue that is the opioid crisis in America. Right now, America faces an average of 115 opioid overdose deaths each day. Over 270 million prescriptions were written in 2012 alone and are a big contributing factor to this issue. As data scientists, we can’t really question whether doctors are pushing prescriptions from sponsorship from pharmaceutical companies, or if they are constantly barraged by patients claiming they are in severe pain and actually need the medication, or really, most other factors for that matter. So, what I wanted to do was try and find the likelihood that a provider would prescribe an opioid and identify any patterns, if any, from that data.

This is important for more than a few reasons. The biggest being, the problem of addiction is a hard thing to tackle. It is an issue I was constantly faced with, involving opioids, all through my 20s after a motorcycle accident. Something I learned from my and through NA meetings was that addicts specifically seek providers that are more likely to prescribe opioids over others. There are message boards, subreddits, social circles and even entire websites dedicated to identifying those providers and publishing the information so that people can more easily have access to opioids.

I started by going to Kaggle and using a dataset from <https://www.kaggle.com/apryor6/us-opiate-prescriptions>, which is a subset of the data from Centers of Medicare and Medicaid Services. There were multiple datasets including overdoses specifically related to opioids, and opioid deaths, the different types of opioids in current circulation by providers, and the provider list of opioid prescribers. After cleaning the data there was some pretty interesting finds. California, was easily the state facing the largest opioid problem with a sum of 11 million deaths over the course of the year. Not surprisingly, there was also a correlation with the amount of prescriptions prescribed, and the total number of opioids prescribed by state. Something that was surprising, though, was that the two highest specialties prescribing opioids were Family Practice and Internal Medicine. I’d like to say that is because those two offices probably see the most traffic. A limitation here is that I didn’t really get to see how many people went into each different specialty which sort of unbalances things here. If we had a balanced version of it, it may lead to better findings.

Another limitation in this dataset is the state of the patient. I’d really like to know if any of the patients were terminally ill, and therefore, put on a terminally ill opioid like Fentanyl. Fentanyl is a drug that is 100 times more powerful than morphine, and usually saved for patients that are specifically terminally ill. Even then, there could have been other factors in the patient’s life that could have led to the death. Were other drugs being used in tandem with the opioids? Benzodiazepines have been known to be mixed with opioids quite often and can lead to your heart stopping. An addict usually tends to abuse more than one drug at a time when opioids are involved, because it can be extremely difficult to keep the supply up while the tolerance within each patient gets bigger. I think with some more specific data, we can really get down to the nitty gritty of what is actually going on here.

Since I was dealing with a classification problem, I decided to use logistic regression, KNN, Decision Trees, Bagged Decision Trees, Random Forests, Adaboost and took a shot at a Keras neural network. My target was whether or not opioids were prescribed by the provider, which increases the likelihood of the patient being given opioids. Multiple models actually had a score of over 90%, and I do not think they were overfitting. Logistic regression, Decision Trees, Bagged Decision Trees, Random Forests all performed extremely well.

My big takeaway here, in a non-technical explanation is: stay away from opioids at all costs, unless ABSOLUTELY necessary. The addiction rate is high, the mortality rate is high (from the exploratory analysis), and the chances of getting prescribed the medication is high.