Saving Lives by Detecting Opioid Prescribers

# Context

Drug overdose and opioid-involved deaths continue to increase in the United States. Opioids are powerful pain relievers that disrupt transmission of pain signals through the nervous system. According to the Centers for Disease Control, more than six out of ten overdose deaths involve an opioid. Unfortunately, the number has been in perpetual increase during the past 15 years (1), highlighting a growing addiction problem. As to avoid systematic opioid prescriptions as pain relievers, it would be interesting to identify unnecessary prescriptions, limiting therefore opioid addiction rates and overdose deaths.

Our aim behind this challenge is to use predictive modelling to identify significant opioid prescriptions. For that we will be using data (2) with prescription records for 250 common opioid and non-opioid drugs written by 25.000 professionals in 2014.

Understood !

# Material and Method

## Presenting the data

The dataset we used is a subset of a larger one, published by (2), containing 25.000 prescription instances. For each instance, we have prescriber-related characteristics like his gender, his specialty and his state of origin. The data also contains 250 numerical variables related to drug names, where for each prescriber, we indicate the number of times his prescriptions included these drugs. Instances are labeled using a binary variable that takes 1 whenever the individual prescribed opiate drugs more

than 10 times in the year. The main data is in *prescriber-info.csv*. There is also *opioids.csv* that contains the names of all opioid drugs included in the data and *overdoses.csv* that contains information on opioid-related drug overdose fatalities.

## Preprocessing the data

As we all know, machine learning is only as effective as the data that drives it. In other words, if you want to implement effective machine learning, you need to pay attention to data quality. For that, we had to preprocess and clean our data before starting modeling.

Preprocessing steps included:

* Modifying Opioid names-formation to match the corresponding attributes in the main data file.
* Cleaning up Categorical variables (identifying aberrant categories, correcting typos and cleaning the abbreviations)
* Factorizing categorical features (like Gender, State, Specialty)
* Eliminating opioid attributes. This step is essential in the context of our objective. Since we aim at detecting opioid prescribers, opioid-related attributes shall be removed, otherwise we would be cheating!

Nothing better than clean well-structured data! We’re good to go!



## Predictive Modeling