Medication Prediction: Sprint02

|  |  |
| --- | --- |
| **Stakeholder** | - |
| **Business contact** |  |
| **Duration** | 1 Day |

# Sprint 02: Developing Specialty Model

## Key findings:

1. Bullet list summary of key findings.

## Sprint Aim:

To research all possible resources that can be used for an extraction and sorting pipeline. Furthermore developing a prototype organising system to prep the data for extraction.

## Sprint Objectives:

1. Develop the Specialty prediction system
2. Have a suitable API to present the data and heat-map

## Run Model

## Method:

1. Import all packages
2. Import Python Model
3. Select only Dentist data from the model
4. Perform a get dummies on gender
5. Set target on Opioid prescriber column
6. Set your X as all data apart from – NPI, Specialty, Credentials, State and the initial target data
7. Set your Y as only the Opioid Prescriber column
8. Load Sklearn and set the XGBOOST classifier parameters
9. Set your X\_train, y\_train, X\_test and y\_teste
10. Run the model
11. Save Model

## Results:

I tested both models ( One created with Python and the other created with Python)

* Using the code form the Lloyds Datathon
  + "for a given specialty, the model returns the most commonly prescribed drugs associated with the variable "opiate.prescriber" = 1
  + This shows the common opiate and non-opiate drugs prescribed.

1. Both models work and output data, however the R data presents more information (feature importance’s)
2. The only questions I have are:
   1. Is the python script missing information, or do I need to add more data to that model to produce a useful output?

R Data



Python Data

