POC of Classification of Health Status from Medical and Prescription Drug Usage

By Eran Schenker

Project Overview

- Background & challenge
- POC Outline (Part 1)
- POC Outline (Part 2)
- Model
- Results
- Intuition for Features Importance
- Possible Post-POC Steps

Background & Challenge

- The Need

  Profiling health status of health insurance members is essential and has several drivers:

  I. Detect members that are not receiving required medical care
  II. Identify unusual care patterns that could convey health risk

- Value Proposition

  I. By addressing the profiling need, the insurer could anticipate issues and thus manage and possibly intervene to change members' outcomes for the better

- Challenge

  I. Different datasets are sometimes channeled from several different providers
  II. Each dataset holds only partial information about a member
  III. Inferring health status of a member by using 2 different datasets is challenging
  IV. How to achieve meaningful results on health status from several datasets

- Tasks Performed

  I. Defined health status of members and defined the level of the analysis (a member per a certain time period) from **medical diagnosis** data
  II. Hypothesized and implemented a model that learns how to classify this health status by using **prescription drug utilization** data

Proof of Concept (PART 1)

Background

- In the claims dataset there are more then 70,000 ICD-10 diagnoses codes which are not hierarchical and therefore could not be used directly for classification purposes
- These codes are taxonomized into 18 “ccs_1_desc” super categories
- These could presumably provide a definition for the **health status**
- However, since they are on the **claim level** and not on the **member level**, members could have thousands of different combinations of super categories (which still won’t be informative enough) so a shift towards a more robust characterization was needed

Rationale

- A generalized and actionable definition of **health status** was chosen - the **wellbeing** of the member.
- Since the 18 groups are not hierarchical (could not provide a simple mapping of a member’s wellbeing) the decision was to build a **risk score for wellbeing** built on the following assumptions:
    1. The **more diverse the diagnosis profile** of a member is (more groups), the higher the risk is for this member’s wellbeing.
    2. The **more total medical records** a member has, the higher the risk for this member’s wellbeing

Proof of Concept (PART 1)

Implementation

- Generate a risk score for each member:

  A weighted average risk score was calculated: **0.75 X the count of unique ccs groups + 0.25 X the count of records**, aggregated per member per year. *

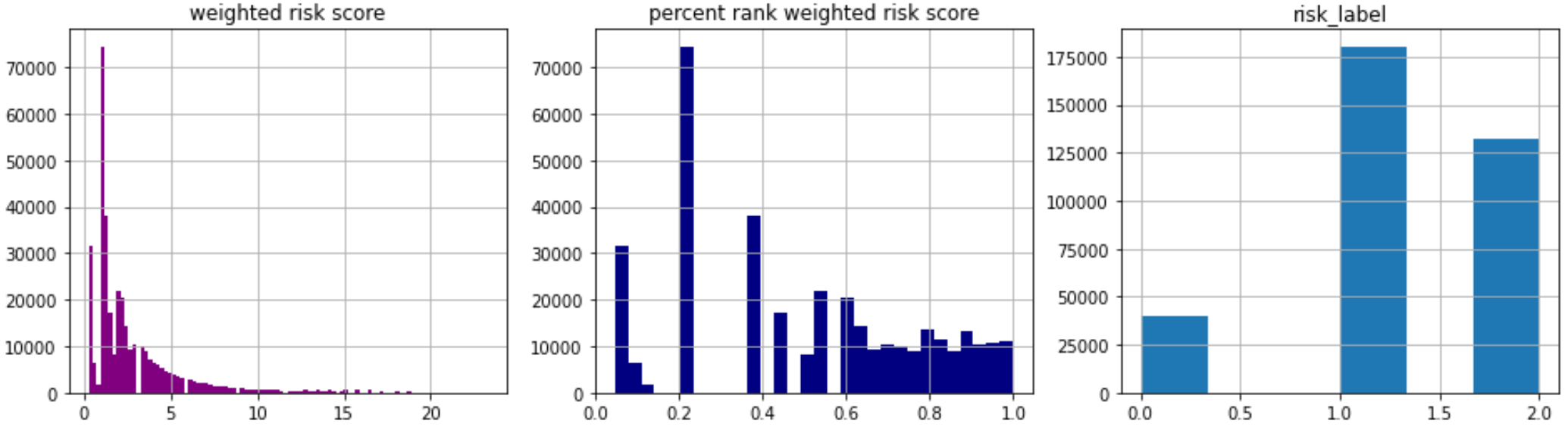
- Create 3 risk score groups (according to risk score):

  Members where classified into 3 groups: low, medium and high risk, according to the percentile of the risk score. The percentile cutoffs were chosen according to table A

table A

Class	Percentile
0	Lower than .33
1	Between .33 and .60
2	Above .60

3 step transformation from **weighted risk_score** to **risk_label**



* Higher weight is given to count of unique ccs groups due to the presumed higher risk that it conveys.
* “ill defined” group and “unknown” group with no ccs mapping were excluded from count of unique groups calculation (used in the risk score)

Proof of Concept (PART 2)

- Goal: Classify health status of members (defined previously as wellbeing i.e. the risk group) by using **prescription drug utilization** data

- The solution proposed: predict the risk group of a member at the beginning of each calendar year for the rest of the year, by looking at the member's data from past calendar years.

- The resolution of the model: The prescription drug data will be summarized (aggregated) by calendar year per member

Proof of Concept (PART 2)

Background

- Prescription drug utilization data consist of more than **20,000 drugs**
- Drugs are taxonomized into **94 drug categories** (highest super category)
- Variables in the dataset are limited "drug categories" "drug_class" and "drug_group", so feature engineering is crucial

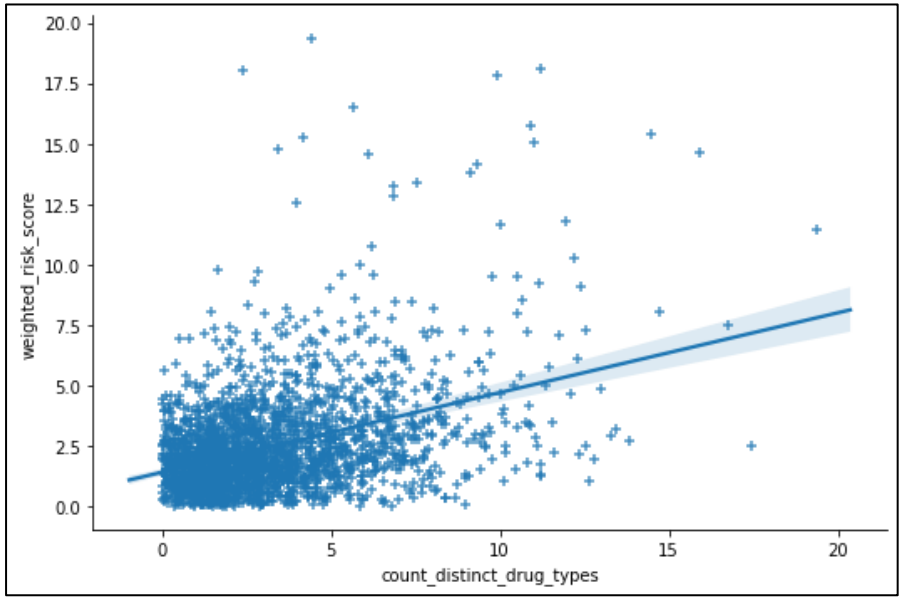
Rationale

- 94 different drug categories and the number of records per member is at the basis of engineering the following features**:**
    1. **"Count_distinct_drug_types"** - **more diverse number of drug categories** a member has (more groups), the higher the likelihood that a member suffers from more than one condition.
    2. **"Count_Prescriptions" -** The **more prescriptions** a member has, the higher is the risk for this member's wellbeing
    3. **94 indicator features** (e.g. "**drug_category_analgesics – opioids"**) indicates if a member had a prescription for a specific category (Some drug categories are more related to severe medical conditions than others)
    4. **Count drug type features –** counts number of prescriptions a member got for a specific drug category
    5. **"count_prescribed_months"** - The total number of months that a member received at least one prescription. Can convey chronic diseases, that may (or may not) be correlated to severity of condition and member's wellbeing
    6. **"STD_monthly_count_prescriptions"** - Low variance of number of prescriptions per month can convey chronic conditions, whereas High Variance can convey acute condition
    7. **"Quaterly_proportion_prescriptions"** – quarterly count of prescriptions divided by total number of prescriptions (for the member). 4 features for every Q. Rational - seasonal peak in Q1 to Q4 in relation to Q2 and Q3 can be correlated to flu related conditions for the member and be negatively correlated to the member's wellbeing

Model

- **Steps and assumptions prior to modeling** –
  - Validating that some features have some correlation to the target variable (figure on the right)
  - We prefer to tune our model for higher recall (than precision) for class 2 due to the cost of falsely classifying a member as “low risk” where they should be classified as “high risk”

- **Modeling steps**
  - Built train set on calendar years 2016, 2017 and test set on 2018
  - Ran classification model (Random Forest Classifier (RFC)) on the dataset to get baseline initial results
  - RFC was chosen due to its robustness with sparse data (due to 94 indicator features), its robust tuning capabilities and relative interpretability (with features importance application)



“count_distinct_drug_types” feature is correlated with the “weighted risk score” which the risk_label is based on

Results

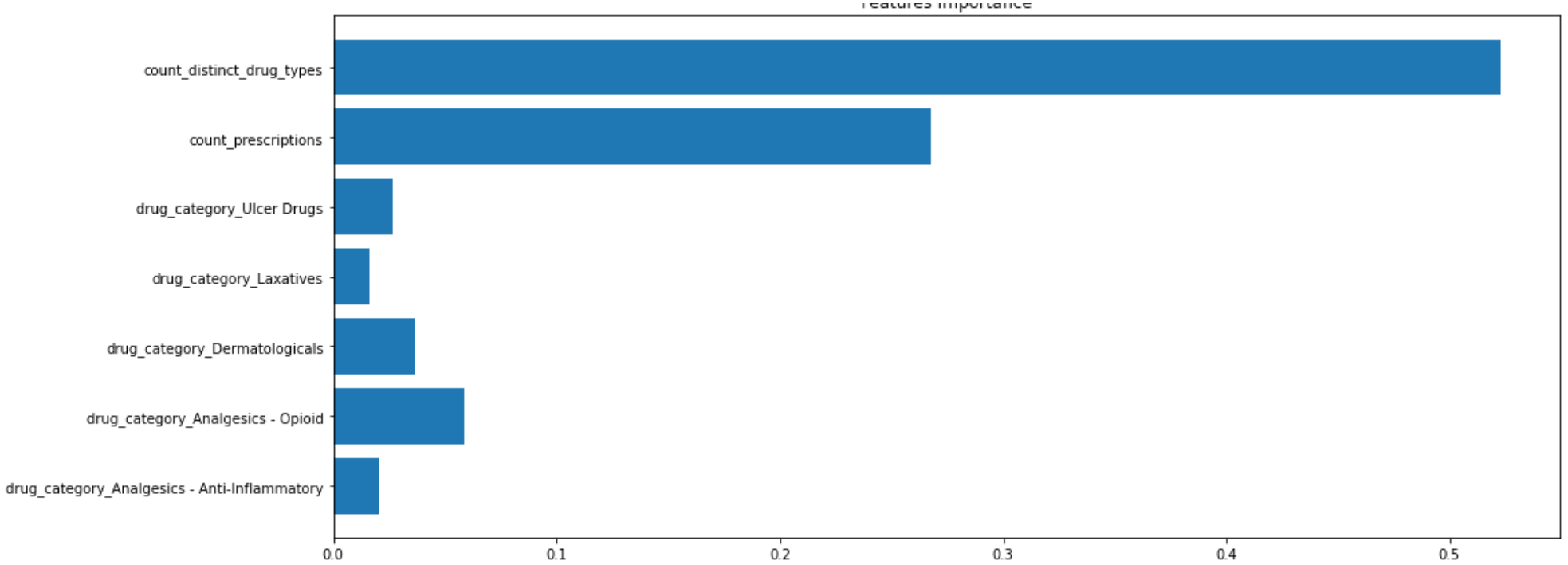
- **Classification Model Results** –
  - Results (on the right) show low accuracy, (class 0 was undetected), yet sensitivity of high risk group (class 2) had the best values (0.72)
  - With this imbalanced performance across the classes, tuning the thresholds to improve class 0 metrics would undoubtedly hinder the metrics of class 2*, therefore other measures should be taken to improve performance

	precision	recall	f1-score	support
0	0.00	0.00	0.00	3995
1	0.70	0.58	0.63	26802
2	0.44	0.72	0.54	13531
accuracy			0.57	44328
macro avg	0.38	0.43	0.39	44328
weighted avg	0.56	0.57	0.55	44328

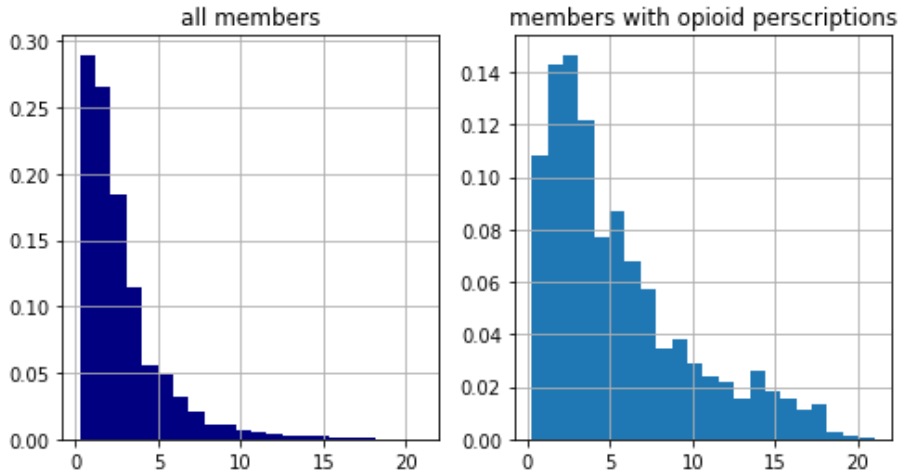
- **Next steps to improve performance of model**
  - Risk score level:
    - Calculate percentiles on risk factors first, and only then calculating weighted avg of risk score
    - Use lower level of ccs categories to refine the risk score (more specific)
    - Implement more assumptions to the risk score (mapping of severity of diagnoses)
    - Label transformation - classify different number of classes (more than just 3)
  - Feature engineering level:
    - Implement features 4-7 (from page 7 yet to be implemented) and add more
  - Consider segmentation of members (e.g. members with chronic conditions) according to prescription activity or drug mapping, then run independent analyses on the different segments
  - Model level – tuning parameters of RFC, feature selection, different classifiers
  - Sampling level – cross validation

* AUC score and ROC were not implemented due to these reasons

Intuition for Features Importance



- 2 dominant features are apparent. These are the engineered features 1 and 2 (discussed on page 7)
- **“Drug_category_analgesics - Opioid”** has the highest importance from the indicator features. An explanation could be that it treats a common symptom - pain that is related to numerous health conditions as could be seen with the shift to higher risk scores for that group (histograms on the right)



Relative frequency of weighted risk score

Possible Post-POC Steps

(Once model performance improved)

- Target risk members found by the model
- Conduct error analysis. The objective is to investigate the misclassified members. The significance in some occasions could be:
  - Identifying misalignment in drug prescriptions due to human error
  - Identifying members who do not utilize their medications
  - Identifying other errors due to misalignment between providers

All of which are “good” errors. By contradicting the model and its assumptions, they drive an opportunity to intervene and change outcomes for members for the better