

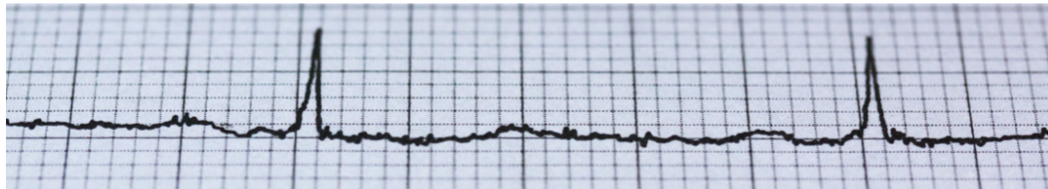


Daniel Shapiro, PhD

Follow

Passionate About Machine Learning R&D and Value Creation. <https://lemay.ai>

Nov 2 · 10 min read



Relapse trigger: Predicting stress with A.I.

From problem formulation to solution deployment, today's article is a comprehensive look at a data science and machine learning journey on the path to help people with addiction caught in the opioid epidemic.

Problem Formulation

[TryCycle Data](https://trycycledata.com) (trycycledata.com) set out in 2017 to develop a data-driven system to help address the opioid epidemic. The early stages of the concept reach all the way back to 2012. Self-help solutions for individuals at risk or in crisis were not working. TryCycle instead took the opposite approach, and focused their technical expertise on finding ways to help health resources stay connected with clients, and monitor their risk of relapse as outpatients. Specifically, TryCycle designed a simple but powerful mobile data collection system, to gather data outside the confines of a controlled care setting. This allows existing clinician-client relationships and treatment to be extended, so that clinicians can better understand and monitor symptoms of relapse before risk levels escalate.

Motivation for studying stress predictors in addicts

The number of drug overdose deaths has never been higher, with the majority of these deaths, 66% in 2016 involving opioids, according to the US Centers for Disease Control (CDC). Improving data quality and tracking trends, [has been recognized by the CDC](#) as critical in helping to understand the extent of the problem, focus resources where they are needed most, and evaluate the success of prevention and response efforts.

Chronic stress is strongly related to drug use, and vulnerability to addiction ([more here](#), [and here](#)). We also know that stress increases addiction risk ([more here](#), [and here](#)). There are ample studies connecting stress to addiction ([more here](#)) and showing that stress causes people in recovery to drop out of treatment ([more here](#)).

Well, we can see that relapse is correlated with stress, but why not simply predict relapse? Why start by predicting stress? Well, predicting relapse is too far down the path to be corrected. We want this tool to predict stress so that the trigger for relapse could be mitigated within the circle of care, well before a relapse is imminent.

Data Collection

TryCycle's unique value proposition has been to focus almost exclusively on the clinician and not the patient. That's because the clinician in their model owns and manages the circle of care. Therefore, the years invested in R&D and development thus far, was built on learning with clinicians in mind. The proprietary dataset used for this article contained over 500 entries from a small sample size of 16 individuals, over a period of five months (May 2018 – September 2018).

The phone location and other data was collected by TryCycle's mobile application along with personal journal entries, and regular self-assessments, as follows.



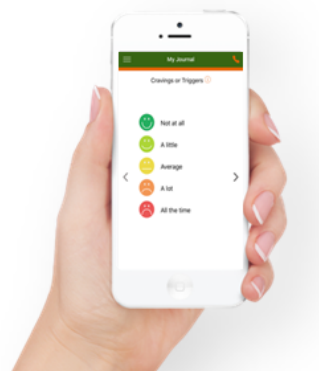
The TryCycle dashboard for reviewing client history and risk of relapse.

Data Collection: Journals

Clients are encouraged to maintain a private journal, which they update on a regular basis. This data is able to be analyzed by machines in a way that preserves privacy, [revealing sentiment and other factors that are helpful in monitoring risk factors](#).

Data Collection: Regular Self-Assessments

Clients are prompted on a prescribed schedule (daily, twice daily, weekly, or twice weekly, depending on how it is set up by the clinic), to complete a technology-based assessment, comprised of eight relapse indicators, also known as relapse cues or warning signs. These factors are related to changes in behavior, attitudes, feelings or thoughts that may precede a relapse. Since many studies show that [stress is a major contributor to the development of addiction and can trigger the relapse process](#), TryCycle included stress as one of their indicator measurements.



One screen in the TryCycle mobile app for collecting self-reports on relapse indicators.

The relapse indicators monitored by TryCycle's assessment are listed below. During the self-assessment process, each indicator is given a score by the client using a series of slider bars that cover the range -1 (bad) to +1 (good).

RELAPSE INDICATORS COLLECTED IN ASSESSMENT
Cravings/Triggers: Strong desire or stimulus to use (your) drug of choice.
Medication Compliance: Are you taking your medication(s) as prescribed?
Fulfilling Responsibilities: Meeting obligations, getting things done, doing the next 'right' thing.
Use of Drug (of choice): Any use of (your) drug of choice.
Mood Swings: Extreme or rapid change in mood.
Stress: A state of mental or emotional strain, tension or pressure.
Relationships that Trigger: Thinking about or interacting with people who trigger use.
HALT: Am I Hungry, Angry, Lonely, Tired?

To understand relapse risk, TryCycle developed a proprietary algorithm that weights all 8 relapse indicators mentioned above, as well as data trends, and variances between calculated submissions, to assign a score and level of risk for each person being monitored.

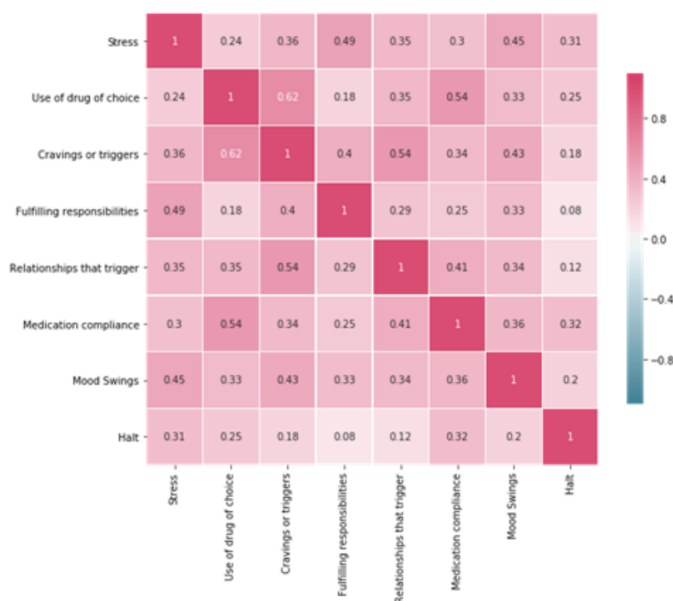
With this data collection system in place, the next question was how to make predictions from this data about future stress levels. Finding correlations between warning signs that precede or lead to relapse was the goal. We decided to focus in on stress, and more specifically to predict stress from other factors in the collected data.

Data Science

Before building a predictive model, we needed to assess and explore the data to obtain an understanding of the features in the dataset. The key disclaimer here is that this dataset was relatively small, as mentioned above, and so these initial analyses are very preliminary at this point.

Investigating How Relapse Indicators Correlate

We examined how relapse indicators correlate with each other. As you can see below there is correlation between every indicator.



Correlation matrix showing how the relapse indicators are all related.

This result makes a lot of sense. These factors should all point to the same thing (relapse). They are all factors in the same risk equation, some stronger than others. We have some criteria for separating correlation into categories according to their strength. Here are our correlation criteria:

Value of r	Strength of Correlation
-1.0 to -0.5 or 1.0 to 0.5	Strong
-0.5 to -0.3 or 0.3 to 0.5	Moderate
-0.3 to -0.1 or 0.1 to 0.3	Weak
-0.1 to 0.1	None or very weak

Using the criteria above, the three strongly correlation indicators were:

1. “Cravings or trigger” and “Use of your Drug of choice”
2. “Medication Compliance” and “Use of your Drug of choice”
3. “Cravings or trigger” and “Relationships that trigger”

Based on these early findings, it may be possible to detect relapse triggers in advance if we monitor the correct indicators. The findings above were a good sign that a predictive model could tell us something useful.

As we dug deeper into the data, and looked at the moderate correlations, we saw a relationship between “Mood Swings”, “Stress” and “Fulfilling responsibilities”.



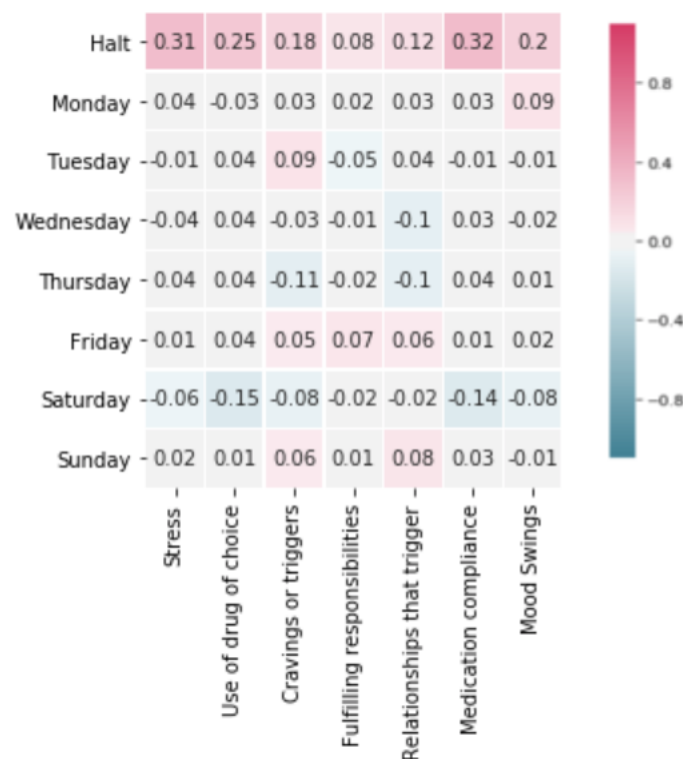
Correlation matrix focusing in on how a subset of indicators are all related.

Looking at the correlation between these indicators we concluded that:

1. "Stress" is correlated with both "Mood Swings" and "Fulfilling responsibilities"
2. "Mood Swings" and "Fulfilling responsibilities" have a moderate correlation with "Cravings and triggers" and to a lesser extent with "Relationships that Trigger"

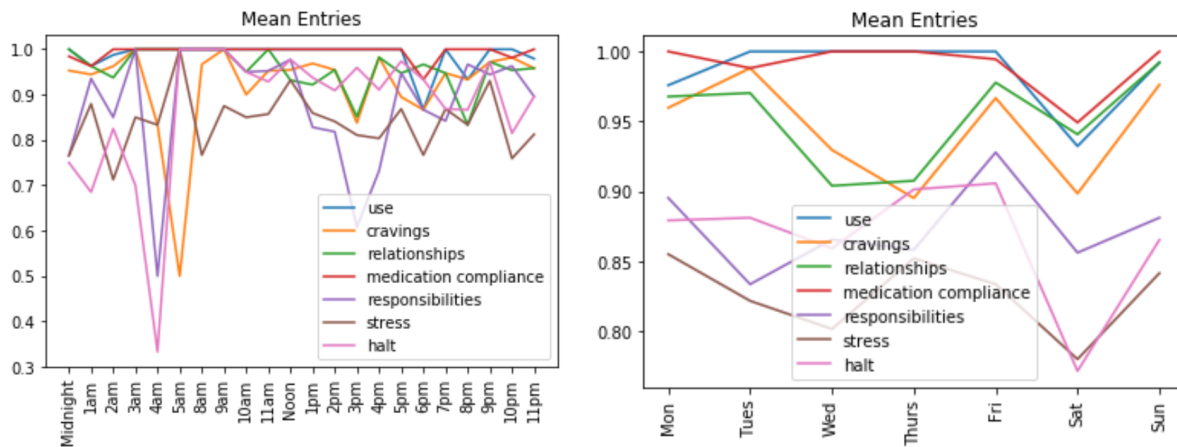
Now, correlation is no causation, but we started to build a thesis around the path from stress indicators to stress itself. We know from the articles mentioned earlier that stress leads to relapse. Perhaps we had a signal here that the factors we collected in our dataset could provide a path to make conclusions. For example: "Stress" leads to "Mood Swings" and "(Not) Fulfilling responsibilities" which lead to "Cravings and triggers" and "Relationships that Trigger" and ultimately a relapse?

We looked for other corroborating information in the data. If we assess the correlation between our relapse indicators against the days of the week, we get the following:



These correlations are all pretty weak. It looks like Mondays may have some effect on mood. That's a confirmation of our intuition that going back to work is not fun. We can also see that "Cravings and Triggers" and "Relationships that Trigger" seem to still match up on the same days (Tuesday, Wednesday and Sundays).

Let's proceed to look at the averages grouped by **day of the week**, and **hour of the day**, to see if there are any clues there (side by side graphs below).



The average value for each relapse indicator was grouped by day of the week (right), and hour of the day (left)

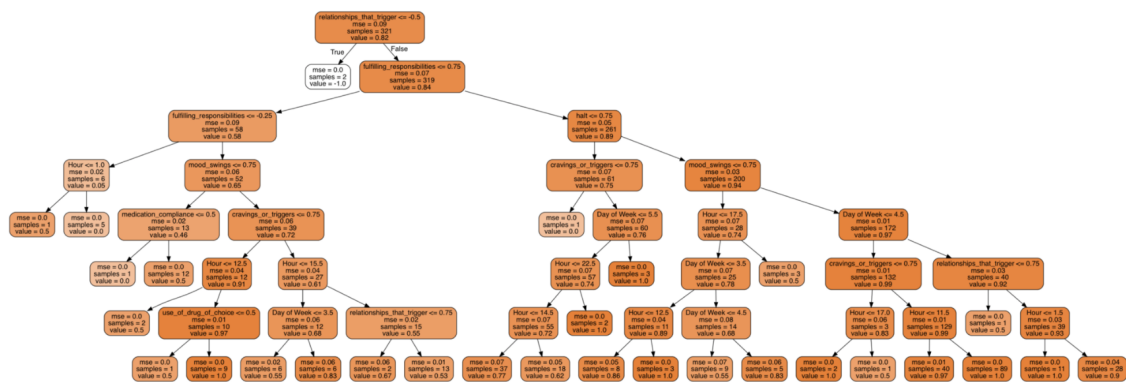
We observed that stress (brown) is consistently one of the lowest values entered if we group the data by hour of day, or day of the week. Recall that low means bad. This tells us that the population reports high stress as we expected. Now, can we predict that stress and then, perhaps, address that stress in counselling or medical visits?

Predictive Model

We want to predict **future** stress levels from the most recently available data reported into the system. This predictive analytical model helps everyone within the circle of care to react before stress turns into relapse.

Digging deeper into the machine learning solution in detail, we went with a [random forest classifier](#) from scikit learn, as this model fit the data well. This type of classifier fits the data by learning decision trees that correspond to the way the data separates. One way to understand this approach is to think of the model as a "[Choose your own adventure](#)" book. Just as the book creates a branching path to go from the beginning of the story to some logical conclusion, this model learns what paths to set out that mimics the patterns in the data. A visual example in the real estate space is [this one by R2D3](#). Scroll down on that link to see a nice animated example of this process. TryCycle's model starts with one relapse risk parameter, and defines a tree of decisions (a.k.a. a decision tree) that says "go left if high" and otherwise "go right". Of course the model also learns what weights to use for each decision.

This data used to make predictions of future stress included the seven indicators and the “Day of week” and “Hour of Day” data for when these data were collected. When the model was fit against the data, we could then make predictions. The resulting decision tree looked like this:



Let’s take a step back and look at the importance of each indicator, to get a better understanding of what the model it does.

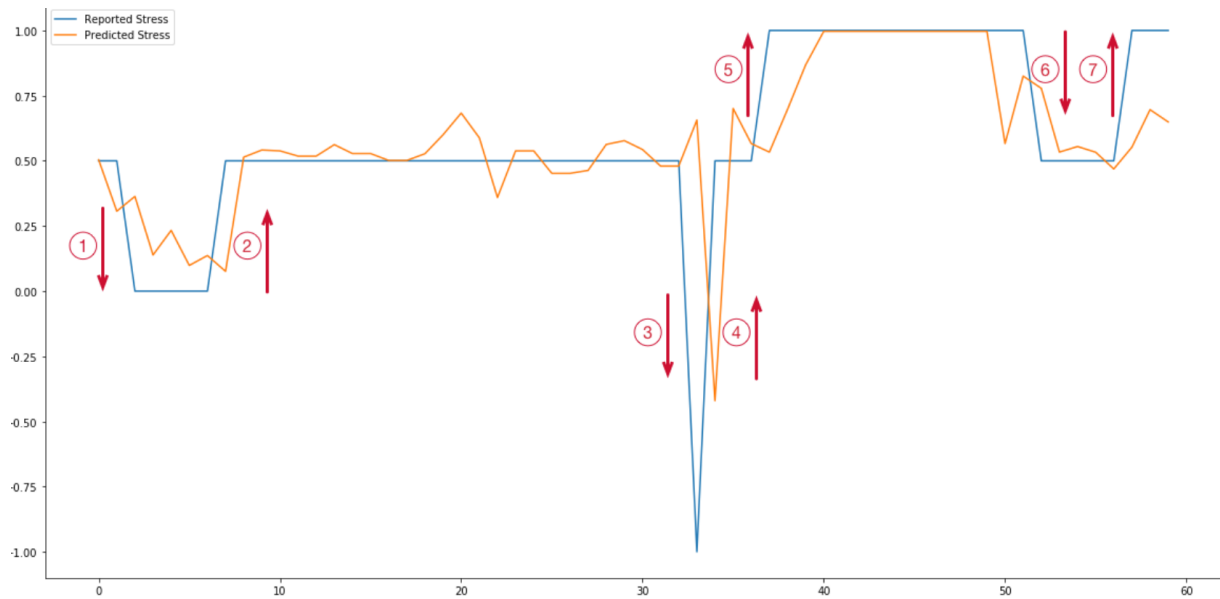
Indicator	Importance
Fulfilling Responsibilities	0.27
Relationships that Trigger	0.2
Hour	0.15
Mood Swings	0.13
Halt	0.1
Day of the Week	0.09
Cravings or Triggers	0.04
Use of Drug of Choice	0.02
Medication Compliance	0.01

Looking at these results, it makes sense that the indicators “Medication Compliance” and “Use of Drug of Choice” would not be good predictors of stress, since we believe stress is their predictor. It seems that responsibilities and relationships are key factors in predicting future stress. The model now checks if the hour is before or after lunch (12.5 hours) or before and after work (14–17 hours), which apparently has a stronger connection to stress than the actual hour of the day. The model also made a distinction between results reported on weekdays and weekends.

Now that we got a sense of how the model makes decisions, how correct are these decisions?

To make the graph below, we pass a indicator data into our model for a given day for one test user, and the model returns a prediction for stress on the next day. We do this over and over again for 61 data points to get a stress prediction curve (orange), and the actual reported stress (blue). Again, the blue line is the stress level reported by this user, while the orange line is the predicted stress value based on data from the day before. If the two lines completely overlapped, then we would have a perfect prediction. We can see that the prediction is actually pretty good. In a few places the stress prediction precedes the actual stress observations!

The accuracy is the average of the difference between each prediction with the next reported value, and the result for this chart above was nominally 95.63% accuracy for this patient. However, the average accuracy of the points is **not** the value we are looking for. What we really want is to be able to predict the **changes** in stress **before they happen**. Looking at the data in the chart below, there are 7 times when the client's reported stress changed.



Observing the prediction of changes in stress level (orange) against the actual changes in reported stress level (blue).

In the above image, we can see that we did guess or were at least trending in the right direction (sometimes even a couple of days in advance) before many of the changes in stress were reported. There are many small changes to the prediction, but the larger changes in predictions were often close to actual reported changes. And so, larger predictions seem to be more precise than the smaller ones, but trending changes can also help with predicting.

We can tune our algorithm to favor either precision (low false positives but sometimes miss stress spikes) or recall (catch more stress spikes but more false positives). The dataset is growing and so the results are expected to improve over time. As more data is collected, we are also thinking about seasonality of the data, and special cases in the data like holidays.

Putting it all together to help real people

TryCycle started off with a problem: identifying relapse triggers and being proactive within the circle of care in response to those triggers. Hopefully, this article gave you a feel for the development and deployment effort that went into the product. The end result of this data science and machine learning journey is a system for helping addicts caught in the opioid epidemic.



A message from TryCycle Data: We are looking to spread our approach to more healthcare organizations. **To book a demo** and understand the benefits of quantitative outpatient care, please [click here to get in touch](#).

This article is about predicting stress in recovering addicts. For a broader overview assessing the tool overall, see the following 2018 case report:

- [Case Report: Assessing TryCycle, a Smartphone-Based Application and Clinician Tool, in Assisting a Client in Recovery from Alcohol and Opioid Abuse](#)