

Case Study

**SCREEN OUT TOXIC
RECIPES USING AI**



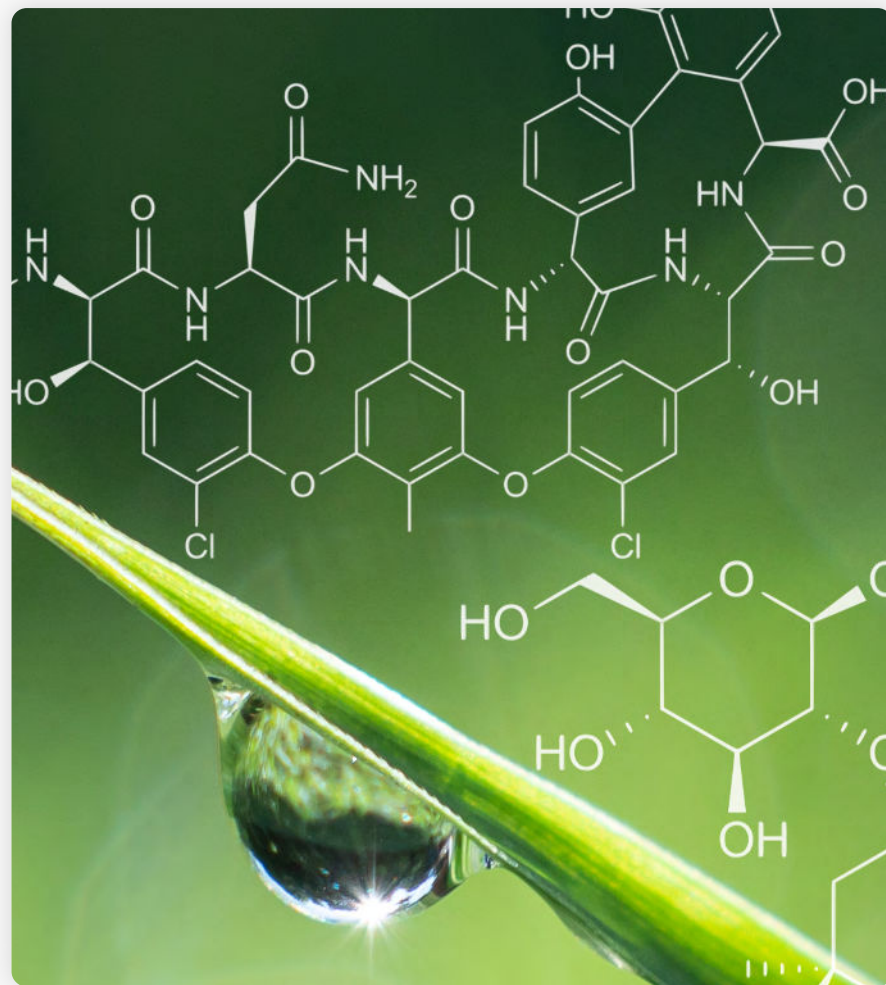
Screen Out Toxic Recipes Using AI

Overview

- Accurate models on complex multiphase formulations
- 4 weeks to develop toxicity model with 82% accuracy
- Screen out toxic recipes before expensive product development
- Reduce animal testing

The Challenge

An innovative customer must ensure that the complex multiphase chemicals they produce are safe to use. While there is robust literature on predictions of toxicity for single chemicals, the challenge here is to screen out toxic formulations early in the development cycle based on their ingredients. This capability would reduce the number of formulations that go through to final animal toxicity testing and then fail, saving time and resources.



The Approach

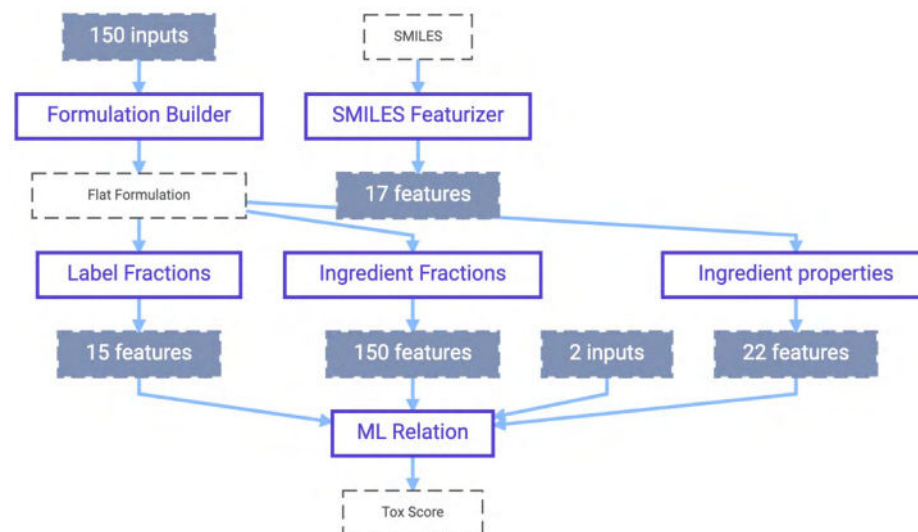
The overall approach was to predict one aspect of toxicity of the final product using knowledge of the ingredients. Data from 36 distinct formulations was used to train the AI model. In total 108 rows of data were used, as most formulations had been tested at 3 different dilutions, a standard part of the test involved. (Diluting a formulation until it no longer causes toxicity problems can also be done as part of product development; however, the product needs to also remain effective at the increased dilution.)

Data available:

- SMILES strings representing chemical formulas for the active ingredients¹
- Measured properties of active ingredients
- Weight fractions of all ingredients
- Ingredients categorized by label (e.g., active, adjuvant...) with their fractions
- Formulation recipe and dilution steps
- Toxicity score (for the aspect being predicted)

SMILES notation is a way to convert a chemical formula into a string of computer-readable information. The Citrine Platform can interpret this string and turn it into a collection of 30+ AI-ready, calculated data points. For example, molecular weight can be calculated from the SMILES notation.

The model



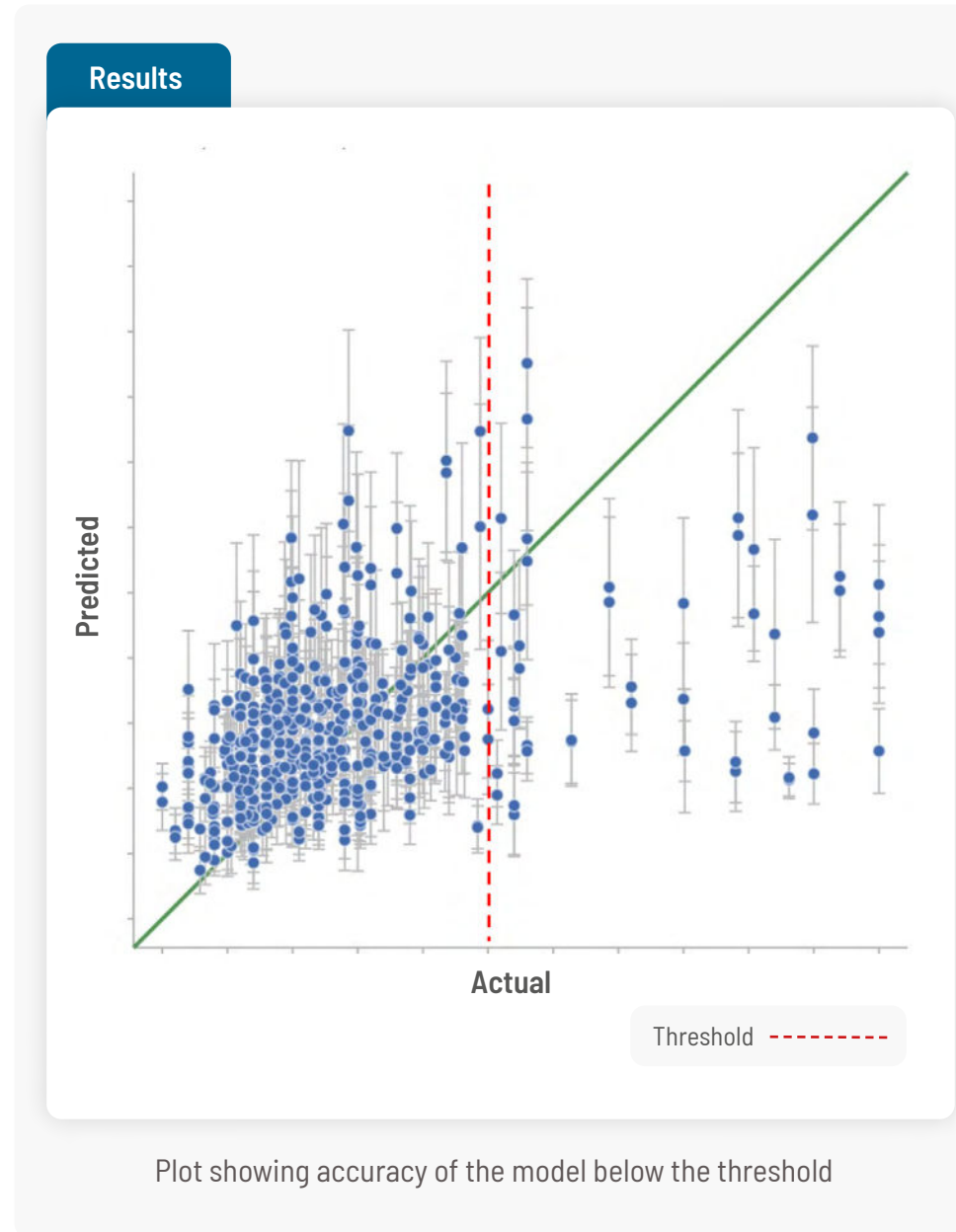
The data was used as laid out in this schematic of the AI model

¹Weininger D (February 1988). "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules". Journal of Chemical Information and Computer Sciences. 28 (1): 31-6. doi:10.1021/ci00057a005.

Results

Cross validation was carried out to calculate the accuracy of the model. While the toxicity data available was on a scale, in reality there is a threshold, and anything above the threshold is a failure. Cross validation showed 82% accuracy in predicting pass or fail for the toxicity test on the formulations in the training set.

The model was then validated on a "hold-out" set of formulations that were not in the original training set. The model predicted the qualitative toxicity result (pass or fail) as well as the quantitative toxicity score correctly 5 out of 6 times. On checking which features the AI model attached most importance to in its predictions, the features did match with the scientist's intuition, making the researchers more trustful of the model.



Plot showing accuracy of the model below the threshold

Benefits and next steps

The customer's researchers were excited by the results achieved in just 4 weeks. This model can already be used to screen out formulations that are likely to be toxic before lengthy product development and expensive testing occurs.

Recipe Screening Model

Ingredient and recipe data

AI model

Predict Toxicity Score

With this model they can also predict toxicity scores on a dilution curve, without further testing.

Final Product Testing Replacement

Ingredient and recipe data

AI model

In vitro test data

AI model

Toxicity Score

The next step in the project is to build a model that can predict the toxicity score based on data from in vitro tests. In vitro tests are cheaper, quicker, and less controversial than animal testing.

Having demonstrated the great value of the Citrine Platform and won the trust of researchers at this company, Citrine is looking forward to seeing the Citrine Platform rolled out to other business units and research teams.



Arrange an online meeting



Listen to our podcasts



Download more case studies



Subscribe to our newsletter

Citrine Informatics Inc.
2629 Broadway St
Redwood City, CA 94063

citrine.io
info@citrine.io

© 2021 Citrine Informatics Inc. All Rights Reserved.