intellegens

CASE STUDY

Developing food formulations at Yili

Executive summary

Global dairy products leader, **Yili**, applied **Alchemite™** machine learning to study UHT whipping cream formulations – a commercially-significant product-line for food services such as large-scale bakery operations. The product must show reliable properties over a nine month shelf life, achieved by using ingredients including stabilisers and emulsifiers, and by controlling processing. The project team built a reliable machine learning model based on 2-3 years of time series formulations data, and showed how this model could be evolved and applied as new data was continuously added. Missing data was imputed and a hierarchical modelling approach was employed to allow earlier months' tests to be used as inputs to improve predictions of later shelf life. A key learning was understanding which ingredients impacted target properties, enabling some additives to be dropped, speeding up the development process.

Intellegens Ltd., The Studio, Chesterton Mill, Cambridge, CB4 3NP, UK

The challenge

Global dairy products leader Yili worked with Intellegens to apply machine learning to food formulation. The project was presented at an Intellegens webinar in December 2022.

The focus of the project was **UHT whipping cream**, a product with a 9 month shelf life used in bakeries and other food services. Performance throughout the lifespan is critical, particularly whipping properties: how long you need to whip it, and the yield: how much air can get into the cream to give the appropriate volume (or 'overrun')? The desired properties are not achievable for a product with this shelf life using natural cream, so formulators add ingredients and play with the process – a complex optimisation problem.

Traditionally, development begins with the knowledge of experts, who propose recipe changes that are tested, first on a small scale, then sometimes at factory scale. Properties are measured and analysed, and adjustments made to the recipe, often with advice from suppliers, as the team iterates around a test-and-improvement cycle. Because of the long lifecycle, new results continue to come in from tests on earlier batches, even as new batches are being made and tested.

Long timescales, scale-up, and many ingredients all make standard 'Design of Experiments' approaches impractical, as too many tests, taking too much time, is needed to cover the design space. Can machine learning guide formulators?

Can machine learning guide formulators in solving this complex optimisation problem?

The project

The team studied formulations created under various homogenisation pressures, containing specified fat and protein content and ingredients from a menu of 9 stabilisers and 15 emulsifiers (some of which were themselves blends). Formulations were subject to two types of storage treatment to understand differences between ideal and real-world environments on product performance. Test data had been collected over 2-3 years, comprising 15,000 data points. The total number of additives labelled on a pack is limited to

2-3 of each type. Test batches meet this constraint, generating sparsity in the dataset. In each row of data, most columns representing additive content are empty.

Outputs, measured at varying storage time intervals for each batch (to a maximum of 9 months), included: physical stability, separation, pouring properties, whipping behaviour, yield, and tests for specific applications. For operational reasons most batches had results missing for some months. Combined with the constraints on additive data, this makes the dataset challenging for classical machine learning algorithms.

Alchemite™ machine learning was used to extract value from a real-world food formulation dataset

The project team applied the Alchemite™ machine learning software [1] to understand this real-world experimental dataset. In the webinar, **Matthias Eisner**, Innovation Manager at Yili, described how the team began with a subset of the data and then gradually expanded the amount of data added to the model. This enabled them to understand how the quality of the machine learning model changed as more data was added, and to see how machine learning could be applied in scenarios where new data was constantly being fed in.

The project aimed to answer three key questions concerning the whipping cream:

- Can machine learning provide insights into product behaviour over time?
- Can it determine which factors do and don't influence the target properties?
- Can it suggest mechanistic insights and cross-correlations (for example, to what extent can behaviour in later months be predicted from earlier results)?

Outcomes

A key issue was the need to handle time series data. During the webinar, Matthias Eisner highlighted the support from the Intellegens Science Team in working out how to integrate time series modelling into the analysis.

"We could quickly drop a number of the ingredients... this was a big learning and helped speed up development"

Matthias Eisner, Yili

The project team built models with good quality metrics (R^2 values) from the data. These were used to predict missing values: both numerical values with their uncertainties, and 'categorical' data, e.g., was a batch likely to separate (Fig 1)? Real performance data from earlier months' tests was used as inputs to predictions for later tests. This hierarchical modelling approach extracted maximum value from the data, enabled insights to be gained earlier, and could in future be used to reduce the amount of testing.

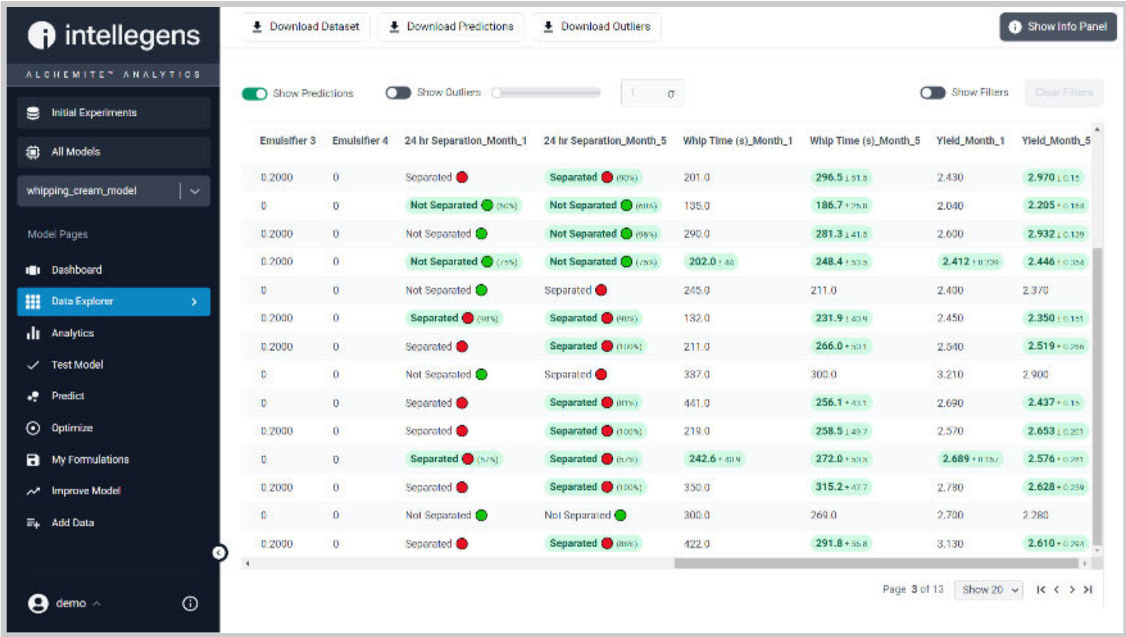

Figure 1. Imputing missing data from the cream formulation dataset in Alchemite™

A particularly useful finding came from the ability to identify which inputs were most significant in influencing the target properties (Fig 2). Matthias Eisner identified this as the most important benefit of the project: "we relatively quickly could drop out a number of the ingredients we had been testing and we didn't continue with them because they did not affect the performance in the way we thought they would, or the supplier claimed. This wasn't obvious if you just looked at them one-by-one, because you always have some cross-interactions. This was a big learning and helped speed up the development."

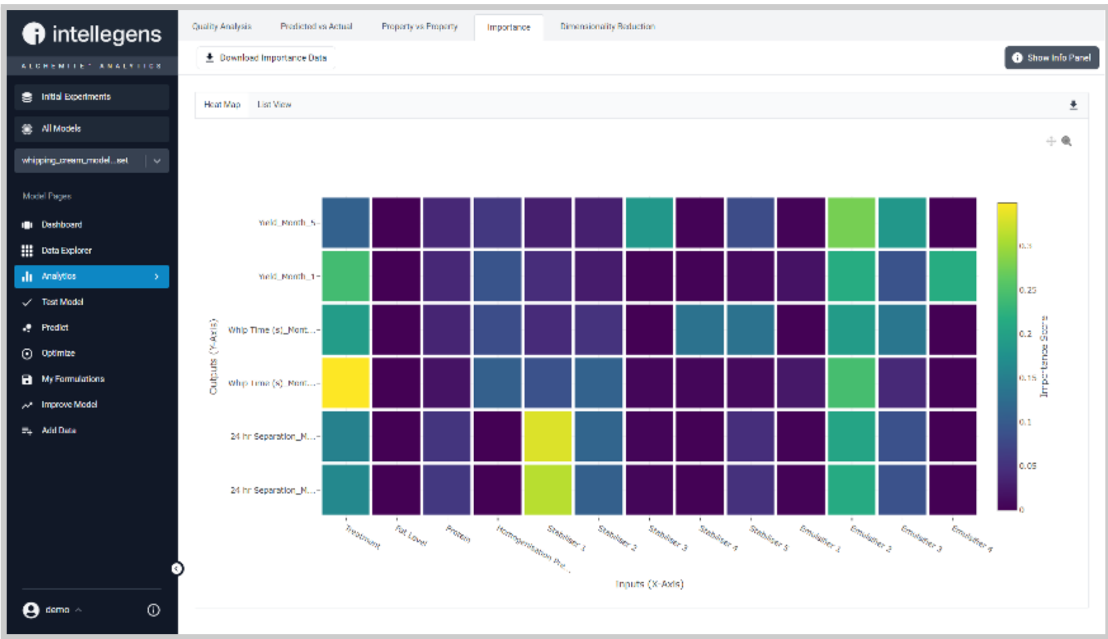

Figure 2. Importance chart shows which ingredients influence target properties.

Summary

Yili and **Intellegens** successfully applied the Alchemite™ machine learning software to the development of cream formulations. In answering the three key project questions:

- Machine learning was able to provide valuable insights into the behaviour of the cream formulations over time.
- It could determine which inputs influence the target properties – this allowed Yili to remove some ingredients from consideration.
- The model was able to identify some useful cross-correlations and to predict likely behaviour in later months from earlier test results.

In addition, the project was able to predict optimised formulations based on a set of input variable constraints and to demonstrate how a machine learning model can be continually evolved and applied as new data becomes available.

About Yili and Intellegens

Yili Group is Asia's largest dairy company and has remained in the leading position for eight consecutive years. Yili retained its ranking among the Top 5 global dairy brands in 2022. The company is highly active across all dairy categories, including liquid milk, milk powder, yogurt, ice cream, cheese, and more. Beyond its core dairy sector, the company is continuously evolving its product portfolio and entering new business lines, such as mineral water, plant-based beverages, and lactic acid bacteria drinks. **yili-innovation.com**

Intellegens, originally a spin-out from the University of Cambridge, provides unique machine learning software, Alchemite™, that is applied to accelerate innovation in industry sectors including chemicals, food and beverage, and materials. **intellegens.com**

References

1. "Alchemite™ deep learning - solving complex problems with real-world data", *Intellegens White Paper* (2021)

www.intellegens.com | info@intellegens.com | @intellegensai

