Data and analytics

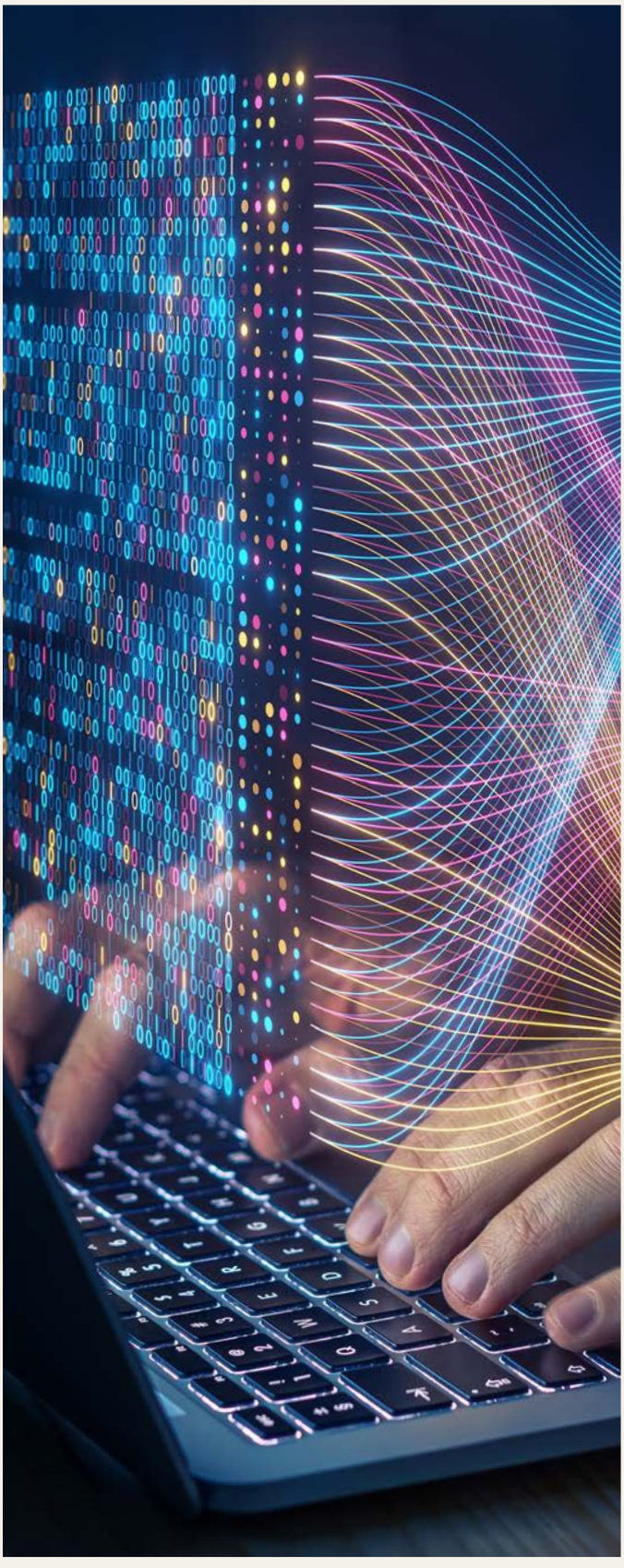
A case study exploring innovative new ways to use data science and artificial intelligence

CRODA

Data analytics

Croda is pioneering the use of data science and artificial intelligence (AI) across its business — from internal operations, to helping customers achieve both commercial and sustainability goals. By exploring innovative new ways to use this technology, Croda has protected existing revenue through targeted customer retention; reduced the amount of land farmers need to grow flourishing crops; and cut the time its scientists need to identify effective formulations for a range of products — from months to days. It is one of very few companies in this sector to investigate the potential applications of advanced analytics to this extent.

In 2019, scientists at the Massachusetts Institute of Technology (MIT) found a new antibiotic, halicin. In May 2023, another team found a second antibiotic, abaucin. What marks these two compounds out is how they were identified. In both cases, the researchers used an AI model to search through millions of candidate compounds to identify those that would work best against antibiotic resistant bacteria.



The challenges

All large businesses generate huge volumes of data. From their customer interactions and the work of the R&D teams, to the operational processes they run every day, the data they generate grows exponentially. The challenge for these businesses is how to harness the value of that data. Drawing real value from its data was one of Croda's major opportunities, but it came with significant challenges.

Firstly, the data itself was not well-organised, it was generated in different forms by many different parts of the business, of which several were past acquisitions and therefore not necessarily running the same systems as the rest of the organisation.

Secondly, expertise in data science was limited and, where it was available, it was dispersed. There was no single corporate-level capability in this field, and no connected effort to generate a return from the data Croda was generating.

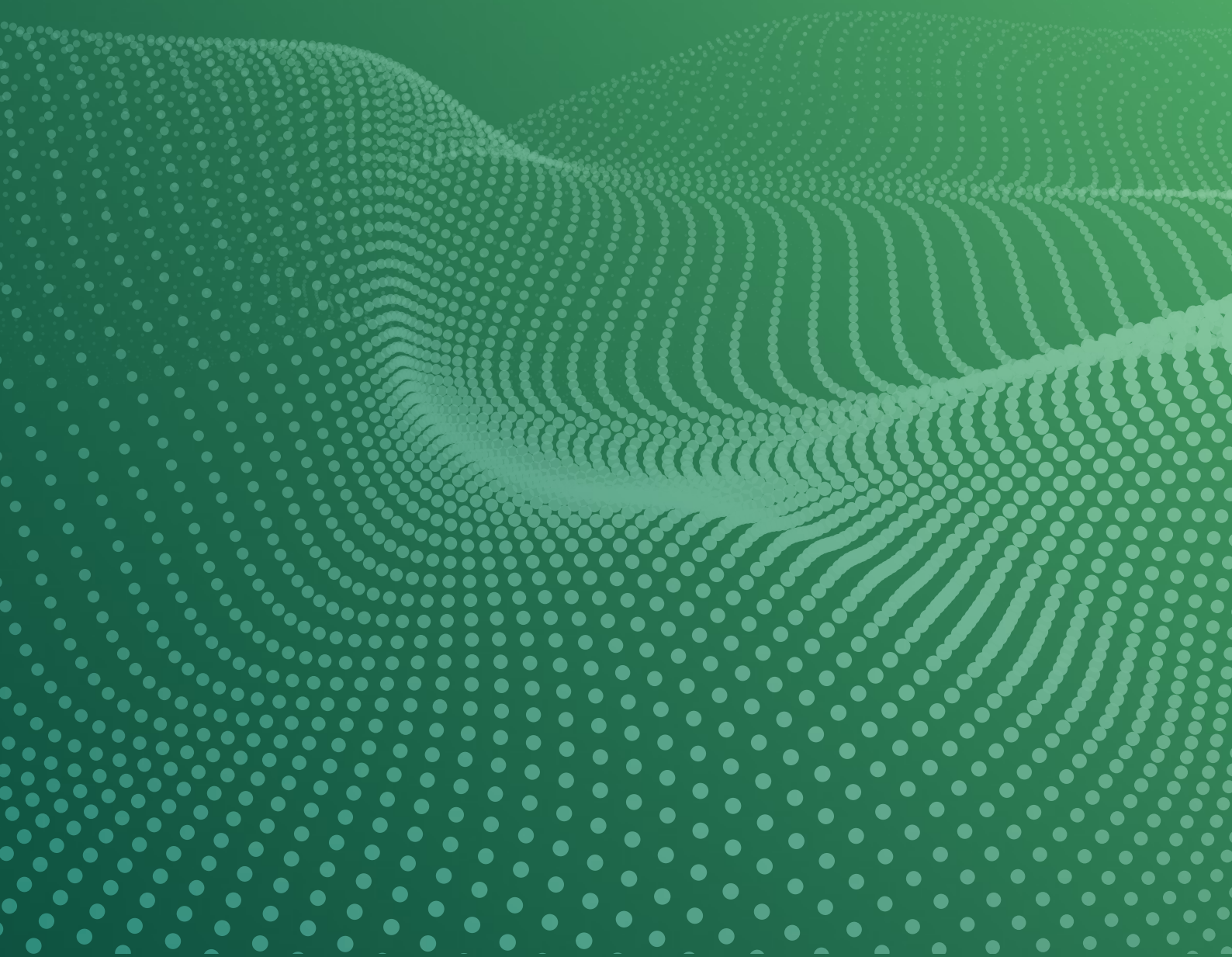
The process and the solution

To address these challenges, Croda made two important changes. Firstly, approximately 10 years ago it implemented Qlik — a data visualisation tool that allowed the company to create data dashboards that made it easier — and faster — to access and use raw data drawn from across the company's systems.

Secondly, on the back of the successful uptake of this tool across the company, Croda established data science as a widespread strategic principle of commercial success and invested in both people and technology to support its use across the company.

It now has capability in data science, AI, machine learning, and computer modelling / simulation. By doing so, Croda has become one of the first companies in the chemicals and ingredients sector to not only possess this expertise in-house, but to launch digital patents in this area.

Three of the people playing a critical role in this shift, but who also demonstrate the eclectic backgrounds at Croda that fuel innovative thinking, are Daniel Massey, Alexander Semenov, and Marie Banks.



Key people involved in data analytics at Croda

Daniel Massey

Data Scientist

Daniel joined Croda 15 years ago having completed a professional plumbing course. He began his Croda career in the maintenance team, later moving to the marketing team where he began to take a keen interest in the value of data outputs from marketing and sales. When the opportunity arose to join the corporate drive towards the increased use of data analytics, he jumped at the chance. Today he works as the interface between the data scientists and those trying to benefit from the use of data science in their part of the business. He facilitates the transition from research to effective application.

Alexander Semenov

Technology specialist

Alexander joined Croda 10 years ago as a researcher. He came with a PhD in Biology and Mathematics, and initially worked on projects with a focus on simulation and modelling. During his time at Croda, Alexander has expanded his remit into many areas of data science and artificial intelligence. This has enabled him to promote and advance the role of data analytics at Croda, and help the company explore new ways to apply it to many aspects of the company's functions and its work with customers.

Marie Banks

IT Strategy Planning and Delivery

Marie joined Croda in 2022 and was hired to lead the growth, management and effective deployment of data analytics capability within the company. Marie's career has always focused on technology and its value return but has led her, in recent years, from senior IT roles to leadership positions focussed on broader, more strategic digital transformation. Her recruitment at Croda is an indication of the strategic investment in this area, and that the commitment the company is making to data analytics encompasses people as well as technology.

The positive impacts

The proactive and innovative approach Croda has taken to data analytics has benefited the business, its customers, and the environment in several ways:

Increasing the yield from sowing tomato seeds

Tomato seeds cost more per kilogram than gold. This places huge pressure on farmers to maximise the number of seeds that germinate and produce a viable crop. Croda x-rays every tomato seed and uses an AI model, trained on 10bn data points and utilising state of the art machine vision techniques comparable to those used in automated driving systems (eg Tesla), to identify those that are unlikely to germinate normally (or at all) based on what the embryo inside looks like.

Any that are identified as defective are withdrawn from seed batches sold to customers. The market standard for successful tomato seed germination is 85%. With this method, Croda customers achieve 95%. Not only is this a significant increase in the output from each investment customers make in their tomato seeds, but it reduces the amount of land they need to achieve these results by up to 10%.

Improving the speed and accuracy of carbon footprinting

Making a fair assessment of a product's carbon footprint can be done in various ways. Outsourcing the process requires days of manual analysis by a third party. Making an assessment in-house usually limits the accuracy of the outcome result. Croda used data analytics and automation to tackle both challenges simultaneously. It removed the need to outsource carbon footprinting by developing its own in-house technology that could deliver the same process, but better. Croda's carbon footprinting application automates many of the elements of the manual process and it allows for testing of ingredients en masse. This has a huge impact on the speed at which they can be assessed — from one ingredient every few days, to more than 20,000 ingredients in just a few months. Accuracy also improves as a result. Whereas Croda was previously limited to offering 'site averages' for the carbon footprint of its ingredients, it can now offer fair assessment for each one individually. Resulting in more informed customers, more accurate results, and a faster route to both.

Accelerating the process of product development

Creating successful new product formulations is a time-consuming process. Thousands of possible formulations must be tested for viability by scientists before a suitable formulation can be identified. Testing 6,000 formulations would take a single scientist more than seven years. Croda has used the power of data analytics to reduce this to weeks.

By training an AI model on the outcomes of more than 180 successful tests, Croda scientists can now simulate the outcome of different formulations in seconds and then tweak the ingredients, amounts, and combinations to indicate which ones are worth taking forward into physical testing. Croda expects this tool to become a fundamental guide that helps its research and development function save wasted time when determining the best formulations for new customer products.

Reducing the chances of losing customers

Croda's team has built a machine learning model that can help it predict which customer contracts may be at risk. By using a range of data already generated by the business, including sales, CRM, and operational information such as order fulfilment, the data analytics team was able to train the model to identify correlations between customers who have maintained a commercial relationship with Croda and those that have not.

This innovative predictive model was trialed for a year in Western Europe for a single market segment. During this period, it correctly identified 70% of Croda's 'at risk' customers and enabled the company to secure more than £2m of revenue by allowing the sales team to engage those customers with the highest probability of moving to an alternative supplier. Extrapolated globally, this figure rises beyond £20m per year.

Making more effective decisions

Whilst rapid access to data has provided tangible time savings on regular tasks, it is generally acknowledged that the greater impact of this work is the (less tangible) benefit of better decision making. This has allowed Croda to understand its business better as it expands globally — and with more data available at the touch of a button than ever before, Croda can now do things which would have previously been impossible. For example, when Croda divested part of its business to Cargill, its data visualisation and analytics capabilities were critical for being able to pull together the information required for a successful transaction.

Conclusion

The growth of data science at Croda is merely the start of a much longer journey. It has achieved a great deal already but there is considerably more to come. Its commitment into 2024 is to really leverage the company's newly-developed capabilities in data science and push the boundaries of what is possible.