



## Why Correlation-Based Machine Learning Leads to Bad Predictions

[info@causaLens.com](mailto:info@causaLens.com)



[www.causaLens.com](http://www.causaLens.com)



# Causal AI

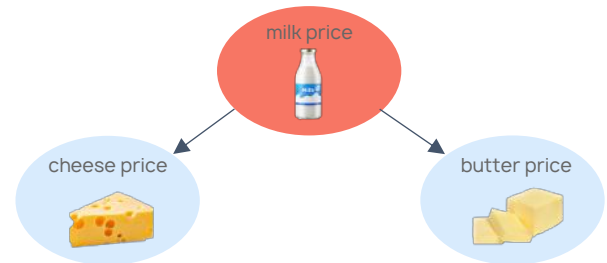
## Why Correlation-Based Machine Learning Leads to Bad Predictions

Machine learning is great at perfectly learning the past. State-of-the-art systems comb through big datasets, identifying subtle historical patterns.

This can be surprisingly powerful when applied to problems for which the environment is unchanging and simple, and the data are plentiful. Flagship examples of machine learning successes involve the constrained, stable worlds of board games and image databases.

However, these machine learning approaches can fail when dealing with messy, real-world data. They often perform remarkably poorly on time-series data types, which are ubiquitous in finance and business.

However, a basic insight -- one which is obvious to us -- is eluding the algorithm. Namely, there is a hidden common cause of both cheese and butter prices: the price of milk. This latent common cause is responsible for the apparent correlation between the two commodities. So, a sudden change in butter prices that has nothing to do with the price of milk will have no effect on cheese prices.



*Milk prices have a causal relationship to cheese and butter prices, which in turn are spuriously correlated*

Unlike machine learning systems, Causal AI does not merely look at correlations. It can autonomously learn the simple causal relationships that seem obvious to us, as well as propose plausible hypotheses about more obscure chains of causality that are less obvious to humans. Because Causal AI is transparent, human experts can partner with the AI, feeding it domain knowledge and real-world context. It does not "overfit" to past data: instead, it is able to zero in on a small number of real predictors. Causal AI learns that the price of butter is not a truly causal signal for the price of cheese, and so is not misled by any change in this spurious correlation.

This example illustrates the pitfalls of making predictions on the back of spurious correlations: these predictions will inevitably fail when the correlations break down.

### Machine learning algorithms perform remarkably poorly on time-series predictions

The key problem current machine learning systems face is that, when it comes to predicting the future, correlations are inadequate. The correlations that have held in the past may simply not continue to hold in the future. Moreover, because correlations are just single numbers, they are not well suited to capturing complex real-world relationships and context.

Let's demonstrate this problem with a simple thought experiment.

### What's that got to do with the price of milk?

Suppose a machine learning algorithm is trying to predict the price of cheese. The algorithm is given access to a dataset with other dairy commodity prices, climatic data and macroeconomic indicators. The algorithm crunches through all this data and identifies butter prices as an important predictor of cheese prices.

Now suppose something out of the ordinary impacts the price of butter. This could be an unusually high inventory (a governmental "butter mountain") or a secular change in consumer behaviour (consumers favouring margarine for health reasons). Following a drop in butter prices, the machine learning algorithm forecasts a drop in the price of cheese.

### When the future doesn't look like the past

What's more, even when machine learning algorithms happen to catch on to the true predictors, they can still end up being badly misled. This can happen due to large-scale catastrophic scenarios, such as the current COVID-19 crisis - dramatic and rapid changes in circumstances without precedent in the data.

Returning to our example: in recent months, dairy prices have been disrupted by unprecedented market behaviour. At the start of the crisis there was a surge in demand for dairy products in supermarkets. This was followed by a slashing of sales as national lockdowns decimated the catering industry.

An algorithm that has happened upon the genuine predictors for cheese prices, including the price of milk, will still be caught off guard by these radically changing market conditions. At junctures in history like the coronavirus pandemic, the patterns that held in the past do not provide much of a clue as to what will come next.

### Causal AI outperforms machine learning under normal conditions, and really pulls ahead in times of crisis

In contrast to traditional machine learning approaches, Causal AI is quicker to adapt to novelty. Causal systems are equipped with "artificial imagination": the ability to simulate events that have never happened, and reason about the hypothetical repercussions of those events. See our [white paper](#) demonstrating how models built with Causal AI adapted to the current crisis three times quicker than state-of-the-art machine learning models.

While Causal AI outperforms machine learning under normal conditions, it really pulls ahead in the kind of extreme circumstances we are seeing in the present crisis.

#### Spilt milk

The costs of poor time-series predictions can be severe. In the context of dairy prices, poor forecasting is responsible for inefficiencies at all stages of the food supply chain.

One way these inefficiencies are felt is in food waste. Sixteen percent of dairy products are lost or discarded globally each year. Waste has intensified as a result of COVID-19, with reports of farmers flooding their fields with millions of litres of unwanted milk. More broadly, according to the UN's Food and Agricultural Organization, global food waste has a combined cost equal to the GDP of France. The financial, social and environmental costs of this are huge.

Improved forecasting could eliminate an estimated 35% of this wastage. Producers and retailers can expect significant return on investment through the avoidance of waste and lost sales, as well as less tangible, but important, reputational benefits. Causal AI can bring about this change by optimising the food supply chain, eliminating waste and increasing efficiency.



### Causal AI actively engages with data: it can simulate interventions and imagine uncharted scenarios

While current machine learning algorithms can passively observe historical correlations, they are unable to distinguish the causal from the spurious ones. As a result, conventional machine learning approaches are, quite literally, stuck in the past -- they are fooled by illusory patterns and are unable to quickly adapt to new conditions. Causal AI has a far more active engagement with data. It can simulate the effects of interventions and imagine uncharted scenarios, just as humans are able to do. As a result, Causal AI makes far more accurate predictions, it is much more reliable, and is more agile in times of crisis.

#### About Us

causaLens is pioneering a completely new approach to time-series prediction. Its Enterprise Platform is used to transform and optimise businesses that need accurate and robust predictions - including significant businesses in Finance, IoT, Energy and Telecoms.

Almost all current machine learning approaches, including AutoML solutions, severely overfit on time-series problems and therefore fail to unlock the true potential of AI for the enterprise. causaLens was founded with the mission to devise Causal AI, which does not overfit, and so provides far more reliable and accurate predictions. The platform also includes capabilities such as autonomous data cleaning and searching, autonomous model discovery and end-to-end streaming productisation.

causaLens is on a mission to build truly intelligent machines that go beyond current machine learning approaches - a curve-fitting exercise. Devising Causal AI has allowed us to teach machines cause and effect for the first time - a major step towards true AI.

causaLens is run by scientists and engineers, the majority holding a PhD in a quantitative field. Contact us on [info@causaLens.com](mailto:info@causaLens.com) or follow us on [LinkedIn](#) and [Twitter](#).