The accuracy of yield forecasts lies in the complexity of the data

Globally, the use of data is becoming increasingly important, enabling many processes to be optimized. At 4cast, physics, computer science, meteorology, and mathematics experts work together with data.

The company uses big data to create precise energy yield forecasts for wind and solar parks. By providing accurate predictions directly to marketers, grid operators, energy service providers, and other industry representatives, they help maximize the potential of weather-driven renewable energy sources. The forecasts can tell you precisely how much energy will be produced in the next 15 minutes, tomorrow, or in the next few days.

The software solutions that 4cast has been developing since 2016 process large amounts of data from different sources to predict energy production as accurately as possible. The company works with machine learning methods and develops deep learning systems that recognize patterns in the data and are thus able to accurately forecast the future, taking into account a wide range of input data.

Here, data science expert Dr Leon Droenner and 4cast Head of Development Dipl. Meteorologist Annekatrin Kirsch give PES an insight into data handling processes and how they manage to draw valuable information from it.

Dr Leon Droenner

Annekatrin Kirsch

PES: It's good to speak to you both. Perhaps we can begin with a simple question, what has 4cast to do with artificial intelligence?

Dr Leon Droenner: The term artificial intelligence is very diverse, and everyone interprets it differently. In the context of 4cast, we are not dealing with an intelligent machine in the sense of a Turing test. The machine and deep learning models we develop only execute what we code, and the predictions are verifiable. Even though the actual models we use are deeply layered and complex, we focus only on the specific task of energy forecasting.

We must engineer and tailor the input data to our needs to achieve the desired precision. It is insufficient to throw raw data at a model and hope to receive a reliable forecast. We must process the data with our domain knowledge and react to different scenarios of input data. Thus, our intellectual work as data scientists uses only the machine's computation power to automatize and unify similar prediction tasks for efficient results.

In particular, we use artificial neural networks for our forecasting. Neural networks and artificial intelligence are often used as synonyms. In fact, artificial neural networks were inspired by biological neural networks based on animal brains, and the complexity, or number of neurons, has grown highly in recent years. However, using the technology of neural networks does not necessarily mean that we are dealing with artificial intelligence. A small comparison: the number of neurons in our use case is comparable to that of an insect's nervous system.

I would prefer to avoid going deeper into the concept of intelligence, because it is much more complex than just the number of neurons. In my opinion, the term artificial intelligence should be used in a less inflationary way. Of course, it is still enormously impressive and amazing what accuracy we can achieve with data-driven models in different areas, in our case, in power predictions.

In general, we need a good amount of data. We want to unravel the versatility of our data and shine some light on the magic that generates our forecasts.

PES: How would you describe the composition of the data?

LD: We distinguish between two different scenarios for which we need data. One scenario is training, and the other is inference, i.e., the actual prediction.

We distinguish between input data, or features, and output data, or target, for training. For inference, we use the input data to make a prediction.

Our input data is mainly composed of numerical weather forecasts (NWP), which physical weather models generate with great computational effort. Typically, we use data derived from different models, such as the German Weather Service (DWD) or the European Centre for Medium-Range Weather Forecasts (ECMWF). The NWP data is generated only every few hours, depending on the model, so we also use live data from measurement stations or satellites. If available, we use live data from plants' sensors as well.

We first need to calibrate our model to the particular site to make a prediction. If no historical data is available, we use site-specific parameters to model the power production based on physical conditions such as the surrounding orography. We call this product the physical model.

If historical data is available, we can build a data-driven prediction, i.e. train a machine-learning model. Meaning, we provide the model with historical input NWP data to match historical production data.

The process can be thought of as a learning process in school. Students make a mistake on a math problem, and through teacher feedback, they know what they need to do differently to solve it. In machine learning terminology, this is called supervised training. We let the model predict based on input data and compare it to historical production data.

Based on the prediction error, we tune the model parameter and check in the next iteration if the prediction is closer to its target until the model can predict the energy production with the best accuracy.

PES: What difficulties have you traditionally encountered in collecting this data?

LD: The biggest challenge is undoubtedly the quality of the target data, that is, the historical production data. As our models learn the relationship between weather and production, the quality of the models depends critically on the quality of the historical production.

Minor inaccuracies in the data have a significant impact. If some timestamps are erroneous, the model runs the risk of learning these errors and repeating them during inference. Examples are shutdowns due to storms, maintenance, bird and bat protection in wind turbines, Icing, and other events.

Another challenge is the quality of the input data itself. We know this from our everyday life. The forecast is for bright sunshine, and while taking a relaxing bike ride, you're caught in an unexpected rain shower. Such inaccuracies in weather forecasts occur more frequently the further into the future the forecast is made. Such is the nature of things since many influencing factors determine the weather.

PES: Have there been other data-related challenges, such as analytics and sharing?

LD: To be able to use data across teams, it must be archived reliably and retrieved easily. In addition, everything in our product must be fully automatable, meaning the data must be systematically categorized and provided fast.

Our data are mainly time series, so we can draw on various methods from the field of computer science and use different databases optimized for the respective application. In addition, we use an interface internally that allows us to read out the required data for analysis purposes in a simple and targeted manner.

A challenge we face are the timestamps we work with. We work with quite a few time series describing the same period. For instance, we make predictions every 15 minutes for 8 hours. It follows that 32 different forecasts describe each timestamp. We encounter the same challenge with the input data, as the weather forecast is regularly updated and corrected. To process this data, you need a complex and systematic data model to keep track of everything.

PES: How does 4cast help to overcome these challenges?

LD: At 4cast, we have experts from different fields, and we have designed specific software for this purpose. Dealing with this complexity is an essential part of our work. Our precise forecasts prove that we have mastered this challenge.

PES: Is this data quite complex and if so, how easy is it to analyze?

LD: Complexity is a subjective term in the eye of the beholder. First, the complexity is due to many factors that can affect the output of wind and solar power plants, including weather conditions, geographic location, plant configuration, and operating conditions. These parameters can be expressed as a number and have a physical dimension. Another complexity is when data is missing or contaminated.

However, these are statistically feasible. Data cleaning is essential for data quality by removing irrelevant and erroneous data. We assess the data base and quality using in-house software that automatically analyzes different time series.

Summarized, the precision of the prediction increases with the quantity of data. A large amount of data is essential for selecting suitable machine-learning models for respective locations. The predictions determine the success of energy trading, stock, and expansion of renewables. This drives 4cast, and with it comes the awareness about our responsibility towards its customers in the quality of its results.

PES: How easy is it for your customers to share the data?

LD: We want it to be easy for our customers to provide us with their data. Thus, there are many ways, and we are eager to find an individual solution for our customers' needs.

For us, it is easiest to get direct access via an API, as these interfaces have become established in software development. However, if this is not possible, there are different ways to access the data, and we adapt to the customer accordingly so that they can continue using their workflow.

Our software automatically reads the data, validates it, and then transfers it to our databases. Together with the customer, we define a schema at the beginning so that the customer does not have to process the data himself. We work in a very solution-oriented manner, adapted to the needs of our customers. 4cast is a service-oriented company, so our customers will benefit from our flexibility.

PES: What are the benefits of all this for the renewable energy sector, and how do you think this will develop in the future?

Annekatrin Kirsch: According to the German government, electricity generation from renewable energies is to be doubled in Germany by 2030. A significant share will therefore be provided by wind and solar power plants. With the 'Wind-an-Land-Gesetz', the wind energy expansion in Germany has to advance significantly faster than before.

By the end of 2032, the federal states must designate 2% of the state area for wind energy. Numerous projects and support programs are accelerating the expansion of renewables. This trend is not limited to Germany. Renewables are being increasingly promoted throughout Europe and even worldwide.

The International Energy Agency (IEA) assumes that by 2030 more than a quarter of the world's primary energy consumption can be covered by renewables. In a study, the international network REN21 shows that renewable energies are developing rapidly and are a significant component of the global energy supply. They provide a secure and climate-friendly energy supply, security of supply, and create development opportunities and jobs.

In this context, precise forecasts are crucial to designing the network utilization as efficiently as possible. Based on our forecasts, power generation by wind and solar power plants can be planned accordingly. Therefore, we make it easier for generators and grid managers to feed the energy provided into the grid and fulfill their responsibilities.

In electricity trading, forecasts that are as accurate as possible are of great importance to keep balancing costs as low as possible. For this purpose, 4cast provides its customers with adapted wind and solar power plant forecasts.

In the future, we will emphasize our forecast quality. In doing so, we must keep an eye on the further development of machine learning algorithms and extend their usage to continue generating fully automated precise forecasts and deliver them in the fastest possible way.

Overall, the accurate prediction of wind turbine yields is one of the critical challenges for the industry and the essential way to improve wind turbine efficiency and profitability. By combining meteorological expertise and data analytics using machine learning techniques, we can intelligently enhance the accuracy of yield forecasts and position the wind power industry for a sustainable future.

www.4-cast.de