

FORECASTING & DATA ANALYSIS PLATFORM FOR GAS AND ELECTRICITY

IMPLEMENTATION OF A CENTRALIZED EUROPE-WIDE FORECASTING AND DATA ANALYSIS PLATFORM

INITIAL SITUATION

Our client provides services in the areas of energy supply, renewable energy, and innovative energy solutions, with a focus on accelerating the transition to a decentralized and sustainable energy future.

The company strives to make the energy supply of the future secure and environmentally friendly through customer orientation, advanced energy solutions, and efficient networks.

Before the migration, our customer used a variety of individually developed forecasting models in each region for short-term forecasts of gas, electricity, and heat consumption as well as wind and solar energy production within Europe.

These models were implemented independently of each other and were largely based on proprietary black box models, which made it difficult to troubleshoot incorrect forecasts.

Much of the source data came from labor-intensive and repetitive manual tasks and used different tools, time granularities, source data types, data quality levels, and standards.

These had to comply with local legal requirements, including customer data anonymization.

In addition, the annual license fees for the previous forecasting tools caused considerable costs.

To reduce license and usage fees, increase data transparency, and meet data governance requirements, a gradual migration of regional units to a centrally maintained and standardized data platform with in-house developed models was planned.

Our consultants were involved in developing the first regional forecasting solution and the associated ETL pipeline right from the start. The models developed in subsequent years were largely based on this initial technical implementation.

PROCEDURE

Part of the success criteria for the project was to individually convince the customer's independently operating regional units of the advantages of the central platform.

After successful pilot phases, the first regions decided to migrate their existing models to the new central platform.

The migration process was approached from the following four key perspectives:

Source Data Migration: The basic processes for sourcing data were changed from manual downloads and Excel-based entries, as well as proprietary third-party tools, to automated API calls and real-time data queries.

The data was migrated from old, decentralized data warehouse solutions to a central data platform and curated.

Project-trained staff ensured that data and model ownership remained with the respective regional units according to predefined responsibilities.

ENERGY

Together with the customer's developers, our consultants laid the foundation for a centrally curated data platform used across Europe.

Thanks to a solution in the Microsoft Azure cloud ecosystem, collaboration between the customer's regional units was significantly simplified. New implementations can be easily realized using a transferable schema. Maintenance and optimization of existing models were drastically simplified.

Additionally, dependencies and costs associated with decentralized, proprietary forecasting and data processing solutions were significantly reduced.

TECHNOLOGIES & METHODS

- Python
- pandas
- Apache Spark
- scikit-learn
- TensorFlow
- Prophet
- Azure Data Lake Storage Gen2
- Azure Machine Learning
- Azure Data Factory
- Azure Databricks
- Azure Functions App
- Snowflake

Transformation and Feature Engineering: Using Azure Databricks, Azure Data Factory, a curated logic component library in Azure Machine Learning, and serverless functions in Azure Functions, we designed a modular, building-block-style solution. Depending on the required execution speed, process complexity, and frequency, the solution is more maintainable, adaptable, easier to understand, and more transparent than the previous setups.

Predictive Modeling: In line with regional requirements, some models from other technological frameworks were reverse-engineered into the internal ecosystem, or completely re-implemented where necessary.

These complex models, some consisting of several hundred features, were analyzed in accordance with data science standards through modeling experiments and exploratory data analysis for feature development and hyperparameter optimization.

Depending on needs and data availability, model complexity ranged from linear regressors to neural networks with hundreds of features.

Distribution: After post-processing, the forecast data was stored in the shared data platform. Weekly training sessions thoroughly prepared analyst teams for using, maintaining, and possibly extending the models.

RESULT

Geopolitical developments are having increasingly noticeable effects on many economic sectors, especially the gas and energy markets. The company's improved responsiveness to unforeseen events now provides a significant competitive advantage and was a major driver for migrating to a common platform with explainable forecasting models.

The centralized and curated data platform greatly increases the flexibility of regional analysts and their collaboration across national borders. It also establishes clear responsibilities in the case of data anomalies or defects. Regular mentoring sessions led by the implementation team enabled effective technical exchanges and rapid, sustainable onboarding to the new platform. The curated AML components follow a low-code approach and can either be used without any coding knowledge or easily adapted using the AML Python SDK. Both approaches were coached during the mentoring sessions depending on the analyst teams' technical capabilities and needs, to ensure sustainable platform usability.

The improved model performance and the go-live of the first of three currently involved regions convinced additional units of the benefits of the centralized solution and the accompanying coaching approach. Over the past three years, the implementation team has grown from about 5 to nearly 30 employees to meet growing demand across the company. Following the project's success, migration to the central data platform is now planned for a fourth and fifth region.

Due to its increasing importance, a central data model for non-forecasting use cases is also being developed, with the goal of establishing a standardized data structure for all European business regions.