it-novum

Guide to Implementing a Predictive Analytics Solution

Contents

What Companies Should Keep in Mind When Implementing a Predictive Analytics Solution

Prerequisite: High-Performance Database Infrastructure

A fundamental prerequisite for efficient data analysis is a high-performance data warehouse or a suitable big data infrastructure that can quickly analyze even large volumes of data. In the context of predictive analytics, it is necessary to aggregate and analyze data from a wide variety of sources. All data sources (machines, processes, products, weather data, etc.) and systems (ERP, CRM, BI, etc.) must be integrated with one another.

Another challenge is comprehensively capturing the wide variety of data types and formats, as well as the quality of the individual data points, and deriving meaningful insights from them. The data should meet all criteria for high quality, such as accuracy, consistency, completeness, timeliness, and uniformity. In particular, for predictive analytics, sufficient historical data must be available so that it can be meaningfully projected into the future using appropriate forecasting techniques.

Selection of Data Sets

After analyzing the business processes and defining the objectives, the next step is selecting the data. Instead of using the entire database, it is usually more cost-effective to tailor the data that will later be analyzed using algorithms and predictive models.

First, the data should be reduced by applying domain knowledge. Next, data that is clearly irrelevant to the use case should be removed. Only then can data that is statistically irrelevant or strongly correlated with other data (a simple example: temperature in Celsius and temperature in Fahrenheit) be excluded from the models.

The selection of data sets always depends on the research question, but can be expanded step by step. Internal data is usually supplemented with external information to obtain meaningful forecasts.

Methodology

There are many different methods and predictive models for deriving forecasts from data. They range from classic data mining methods such as clustering or regression analysis to elements of game theory and machine learning. In the latter case, algorithms are trained to learn from the available data, independently generate a data model, and use it for forecasts or decisions.

Forecasts are only as good as the underlying data. A quality-assured, robust data foundation is therefore crucial. Especially in supervised learning—where the master data forms the basis for the algorithm's learning—a data foundation prepared to a high standard through master data management can yield significantly achieve significantly better predictions than with machine learning using an unprocessed dataset.

The following approach is recommended

The Six Steps of Predictive Analytics

1. Identify data. Potentially valuable data is often located in multiple hard-toaccess places, both internally (data silos in enterprise applications) and externally (social media, government data, and other public or licensed data sources). Data visualization tools can help examine data from various sources to determine which might be relevant to a predictive analytics project.
2. Prepare data. This step is a key challenge in predictive analytics. Many users of predictive analytics spend more than three-quarters of their time preparing the data: calculating aggregated fields, removing extraneous characters, filling in missing data, or merging multiple data sources.
3. Building a predictive model. Predictive analytics tools include dozens of different statistical and machine learning algorithms that data scientists can choose from to build the best predictive model. Which algorithms are most suitable depends on the nature and completeness of the data and the type of prediction desired. Analysts first perform the analysis on a subset of the data, known as training data, and set aside test data that they intend to use to evaluate the model.
4. Evaluate the model. Predictive analytics is not about absolutes, but about probabilities. Models are evaluated using metrics that depend heavily on the underlying use case. To assess predictive power, data analysts use the model to predict the test dataset. If the predictive model can accurately predict the test dataset, it is a candidate for deployment.
5. Deliver actionable insights. A prediction is of little value if it does not enable you to seize a predictive opportunity or avoid a negative event. Stakeholders must learn to rely on model predictions, and those who build the models must learn from their partners across the organization what insights are most actionable.
6. Improve the model's effectiveness. Predictive models are only as accurate as the data fed into them. Over time, their effectiveness may decline or improve. To assess a model's effectiveness and value, newly collected data must be run through the existing algorithms. If the model becomes less accurate, the data scientist must adjust the model (e.g., by changing parameters in the algorithms) and/or seek out and feed in additional data.

Successfully Translating the Gained Insights into Business Practice

Given the complexity of predictive analytics projects, which involve many variables and influencing factors, we recommend proceeding step by step and starting with a small pilot project that is continuously refined as part of an ongoing process. There is always a risk that the predictive models will not deliver the expected results. This can be due to a wide variety of reasons; for example, it may be due to poor data quality. Therefore, it is important to develop the project step by step and gradually expand the data sources.

Good Forecast, Bad Forecast

A comparison of the forecast with actual results reveals whether all the effort put into the respective predictive model was worthwhile. Another measure of a forecast's quality would be a comparison with methods previously used within the company (such as regression analysis combined with a manager's gut feeling) or competing algorithms. The statistical evaluation of the forecasts and the correct metric for forecast quality are important components of a project. If the forecast proves to be valid, the next step is to apply the results to business practice so that they generate added value.

Conclusion

There is no single "right" predictive analytics method for all companies. Which method and algorithm work best depends on the specific goal and the problem at hand. Predictive analytics is not a one-time process; companies must repeat their analysis using new data to ensure that the models are still effective and to be able to respond to changes in customer preferences and competitors. Many companies analyze the data weekly or even continuously

Why you should get in touch with it-novum...

We profitably implement these Business Intelligence and Big Data advantages for companies:

- 360-degree view of your customers
- Departments evaluate their own Big Data themselves thanks to Self-Service Analytics
- Identification of new revenue sources through the intelligent use of company data
- Cost savings by using a data warehouse
- Avoid the costly and error-prone spreadsheet chaos

If you'd like to put these advantages to work in your company too, we should get to know each other!

it-novum offers comprehensive services for your Big Data analytics project:

- Data engineering
- Implementation of Data Warehouses and Data Lakes
- Pentaho/SAP Connector for processing SAP data
- Pentaho/HVA Connector for video stream analysis
- Predictive analytics and machine learning
- Dashboards and data visualisation
- Embedded analytics

Leaders in open-source solutions and consulting for businesses

it-novum is the leading IT consulting company in the German-speaking market for open source-based business solutions. From our headquarters in Fulda and our offices in Vienna and Thalwil, we serve medium-sized and large companies as well as clients in the public sector.

As a Hitachi Vantara Preferred Partner for Big Data Insights and IoT, we are experts in the use of Pentaho. Using our expertise in consulting, training and support, we help companies gain insights from their data and successfully design their data-driven projects.

Your contact

Christopher Keller

Director Big Data Analytics & IoT

+49 661 103 724

christopher.keller@it-novum.com

it-novum GmbH Germany
Headquarters Fulda: Edelzeller Straße 44 • 36043 Fulda • Phone +49 661 103 434

it-novum Branch Offices Austria
Ausstellungsstraße 50 / Access C • 1020 Vienna • Phone +43 1 205 774 1041

it-novum Sales Switzerland
Seestraße 97 • 8800 Thalwil • Phone +41 44 722 75 55