Qlik

A Practical Guide

Automating Machine Learning in Analytics

What is machine learning?

Across many industries, the amount of data available to businesses is increasing exponentially. But on its own, greater volume doesn't always mean greater value. The competitive advantage lies in recognizing hidden patterns in the data to inform action. That's where machine learning comes in — the science of programming computers to automatically find these patterns.

Instead of 'teaching' a computer program to execute a specific task by hard coding explicit rules, machine learning requires a different approach. By exposing a computer to a dataset, machine learning enables it to learn how to do a task on its own by identifying correlations and relationships between data points.

Machine learning is being used for a variety of business applications, from natural language processing to voice recognition and recommendation engines. Taking a step back, its key role is to analyze large sets of historical data and make decisions and predictions based on the patterns the model finds.

Machine learning in action

To illustrate machine learning in action, let's consider a SaaS company selling monthly subscriptions to an online platform and its role in evaluating customer behavior. For example, machine learning can help the business understand the patterns and drivers affecting historical customer attrition, using those same patterns to predict which current customers have the highest risk of leaving.

Example Depending on the customer data collected, the algorithm could yield insights such as:

Customers over

70+

rarely churn during their first year

Female Customers age

40s

on a 'Family' plan rarely churn during their first year

Male Customers age

30s

on a 'Personal' plan are likely to churn during their first year, if they logged in fewer than 20 times during month one

Machine Learning (ML)

ML excels at finding patterns. It can discover significantly more complex patterns in very large datasets to predict how certain attributes will impact the target metric(s). In the example above, this means forecasting future customer churn.

AutoML

As its name suggests, AutoML automates the technically challenging process of building a variety of models based on ML algorithms. This allows people to specify a target metric and find key patterns in a dataset that are often impossible for humans to visualize or detect. Teams can also refine the model, using it to make predictions for both forward-looking data and scenario planning.

The AutoML difference

The Problem

Without an automated machine learning platform, data scientists are typically responsible for the entire process of applying models in a business context. This includes understanding which class of algorithm to use, developing the software and code, training the algorithm with the relevant dataset and evaluating its performance based on statistical analysis. For a data scientist, this is a time-consuming, labor-intensive process which requires deep expertise.

The Solution

Automated machine learning helps organizations use the limited bandwidth of their data scientists more effectively, complementing and accelerating the types and number of problems they can get to. By automating the machine learning process, analytics users and teams can apply models to a variety of everyday challenges. These 'citizen data scientists' are still responsible for asking the right questions, providing an appropriate dataset to answer those questions, and training the model.

Three questions for success

To democratize the application of machine learning models among these 'citizen data scientists', it's important they can answer three key questions about the project in question. The examples blocks below refer to the customer churn example outlined previously.

1 What is your use case?

Example

Our use case is predicting customer churn, and more specifically, predicting which current customers are likely to churn in a defined period.

This first question is typically the easiest to answer. State want you want to accomplish and the associated value proposition. What is the strategic imperative or KPI? How does it impact revenue, cost savings, or risk? Common use cases include predicting customer lifetime value, scoring sales leads, and more — spanning industries and functional areas in an organization.

2 How do you define the metric precisely?

Example

We defined customer churn to mean that a customer signed up for membership to the online platform and canceled that subscription within a year from the date the person signed up. We are looking to predict first-year churn.

Here, it's important to be specific. In the example above, we looked at two types of data — demographic data and first-month usage habits. For AutoML to predict if a new customer will churn in the first year, it needs to know these data points about the customer. A good rule of thumb is to incorporate relevant time frames, specifying both how time comes into play in defining the target metric (customer churn in the first year)

and when predictions will be made in a customer's lifetime about them (after gathering data on the customer's first-month activity).

It's important to use domain knowledge of the specific industry, and problem in question, to select the time periods that make sense. For example, it might make sense to monitor customer behavior for only a week and, based on that, predict whether they are likely to churn within three months.

3 What will your dataset look like?

Customer ID	Gender	Age	Zip Code	Plan Type	Logins 1st /Mo	Avg Mins Logged 1st /Mo	Chrun 1st /Yr
2364275	M	32	11390	Personal	5	5.2	Yes
9568764	F	20	44946	Family	6	7.9	No
4756289	M	65	44856	Personal	10	33.8	Yes
	F	15	15486	Personal	2	28.1	Yes

Example

Description of the rows in the dataset — each row represents a unique customer (i.e., the correct granularity of the dataset is to aggregate it at the customer level).

Description of the columns in the dataset — each column represents an attribute of each customer (i.e., gender, age, logins during the first month).

The final question is simply a matter of describing the rows and columns of the dataset that your team will feed into AutoML to support the parameters set out in the previous question. In this example, if a customer is less than a year old and has not churned, that person cannot be part of the dataset that the model learns from.

That's because it's impossible to determine whether that customer churned in their first year. Whether or not a customer churned after their first year is also irrelevant to this question, since the example is only looking at first-year churn.

What's Next?

From machine learning to business application

Equipped with the machine learning question and having answered the three business questions, you are ready to apply the solution in practice. With the right preparation, you will begin to see its value in the organization.

Machine learning algorithms take the patterns identified in your data and make educated guesses to forecast the future. In the referenced example, knowing the probability that a customer will churn within a certain time frame delivers much-needed insights about their behavior. Once you learn which customers will likely churn, you can employ customer retention strategies and begin to ask questions about more valuable customers' purchasing patterns, demographics, and other characteristics that could help keep them on side.

The knowledge gained by empowering analytics teams with machine learning allowing them to find solutions to everyday problems can directly improve business decisions related to your company's long-term growth strategy and financial performance.

Learn more about Qlik AutoML

Bring the power of machine learning to your analytics teams with the simple, code-free user experience inside of Qlik Cloud. Easily create ML experiments, identify key drivers in your data, and train models. Make future predictions with full explainability data so you can see not only what might happen, but why. And quickly publish the data or directly integrate models into Qlik Sense® apps for fully interactive analysis and what-if scenario planning.

Request a demo at

Qlik.com/AutoML

About Qlik

Qlik, with the recent addition of Talend®, delivers an industry-leading portfolio of solutions for data integration, data quality, and analytics. This includes advancements in real-time data, AI, ML, and automation. The most successful organizations are investing in data to make sense of the increasing amounts and varieties of data from diverse sources. The challenge is to effectively integrate, analyze, and act on the data while ensuring its trustworthiness. With more than 40,000 active customers in over 100 countries, Qlik's solutions work with virtually any data source, target, architecture, or methodology to ensure customers have the data they need, whenever they need it.

Learn more at www.qlik.com

Qlik