



Operationalizing Machine Learning at Comcast

Case Study

Summary

Comcast is using immense amounts of data from tens of millions of customers and hundreds of millions of networked devices to deliver personalized content to diverse TV audiences across the US, improve customer care, build resiliency into its products and reduce truck rolls for the service technicians.

The Challenge

With large-scale systems generating massive amounts of data, Comcast needed a solution to run models on complete production datasets - as opposed to relying on sampling - to improve accuracy of their predictive analytics. "We have hard problems to solve," says Drew Leamon, Director of Engineering Analysis at Comcast. "H2O.ai Machine Learning is one of the cooler things that we are working with in order to help us solve problems in a way that's really valuable for the organization."

"H2O enables us to operationalize more easily. As we train a model, we can export it as a Java code or we can use H2O instances as a web service to build some intelligent apps based on our models."

– Chushi Ren, Data Scientist at Comcast

The Solution

Increased Efficiency and Cost Savings by Preventing Avoidable Truck Rolls

Among the primary use cases for machine learning at Comcast are avoidable truck roll models(ATR). Typically, when the customer experiences an issue – such as dropped connectivity or problems with their TV content, they first try to reach Comcast by phone. The service representative attempts to troubleshoot the issue by following a series of steps outlined in a care protocol. If the problem cannot be solved over the phone, the agent schedules a truck roll – an appointment at a customer's home or business.

By reviewing historical data, Comcast noticed that a portion of truck rolls could be avoided by using simple fixes, such as changing the batteries in a remote control, resetting a modem or changing the customer's subscription or preferences.

With the help of H2O, the Comcast customer care team started to build a predictive model to prevent avoidable truck rolls. For the first steps that involve data engineering for large amounts of data, the Comcast team uses Datameer – a tool that can load data from various sources and present it in a spreadsheet-like interface for filtering and organizing. When the data

Industry

Comcast is a multinational mass media company with multiple lines of business - from high-speed Internet, to feature and TV film productions, cable channels, IP telephony, and home security and automation services.

Challenges

- Building models on complete production datasets to improve accuracy of predictive analytics
- Operationalizing and scaling data science to match the company's volume of operations

Solution

- Reliable predictive models to prevent avoidable truck rolls
- Using analytics to improve customer experience through personalized TV viewing suggestions
- Reducing the effort by the customer and customer service through intelligent problem-solving systems
- Using data analysis to measure and improve customer experience

Results

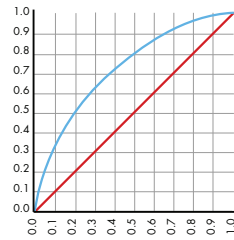
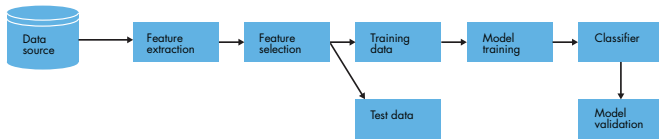
- Cost savings and efficiency gains in reducing avoidable truck rolls
- Accurate predictions of trending content to offer personalized suggestions and browsing options to TV audiences
- Improved metrics to evaluate customer experience
- Intelligent systems to solve problems and improve product resiliency

is ready, it's pushed to H2O and split into training and test datasets. Once the training is completed, the data is validated using the testing dataset. To improve the accuracy of the predictions, the team is continuously fine-tuning the model. For instance, the fact that most truck rolls cannot be prevented, skews the model to favor the unavoidable scenarios. To combat this, Comcast uses the balanced class option in H2O to equalize the samples, along with the subsample method to split the data into more balanced groups before assembling it back together.

Improving Customer Experience Through Smart, Personalized TV Viewing Options

Another big user of predictive analytics is the video research team. Despite many options available to today's audiences, live TV remains the principal way that people consume video content. Studies show that up to half of

Avoidable Truck Roll Machine Learning Pipeline



our spare time is being spent watching live television, making the task of tailoring live content to the needs of different audience much more pressing and complex. Comcast video researchers look at a variety of data, including video assets, subscription information, channel lineups and content purchases – bringing all the data into the analytical platform to develop features around content, browsing options and recommendations. Based on detailed analysis of viewers' behaviors and preferences, Comcast can offer personalized platforms and smarter menus to deliver a better TV experience to their customers.

Using gradient-boosted decision trees, Comcast can reliably predict the popularity of a particular TV show or film 24 hours in advance and make recommendations to the viewers by showing them what's trending before the feature actually airs. To accomplish this, Comcast combines historical data with real-time streaming elements. Users are divided into clusters, while the algorithm looks into what's trending within each of those clusters to produce real-time streaming recommendations.

Using Data Analysis to Measure and Improve Customer Experience

A typical architecture of a Comcast network involves a customer connecting through their cable modem to the Cable Modem Termination System (CMTS), and through CMTS connecting to the Internet. CMTS has multiple ports, which are grouped logically into service groups. Utilization of these groups is currently being used as a customer experience metric, and although there is a correlation between the two, Comcast is looking for a better way to measure and understand the customer experience and prioritize hardware deployment. "We are using the technology to develop a customer experience metric that can be computed across our

entire footprint," explains Leamon. "If a customer has poor video streaming experience, Comcast can alleviate the issue by deploying additional hardware to increase network capacity at that location."

Comcast built a solution to gather and correlate many types of data that are relevant to customer experience across different datasets, aggregate and clean the data, and finally use the clustering algorithm to assemble the data to form customer experience groups.

Evolution of Resiliency: Reducing the Effort by the Customer and Customer Service by Building Intelligent Systems

Additionally, Comcast uses data analysis to make more resilient and reliable products.

In the traditional service scenario, when a system generates an error, the path to resolving it includes the customer contacting the provider, and a customer service agent taking the time to understand the problem, finding the solution and then implementing the solution.

"We wanted to build a system where errors go directly into intelligent online system that leverages Machine Learning to diagnose problems and suggest fixes," explains Chushi Ren, Data Scientist at Comcast. The team designed a system that feeds real-time streams of data coming in from the source application into the data platform, which in turn talks to the rules engine. "The rules engine is powered by machine learning, which is constantly working on scoring incoming data, evaluating models and making decisions on how to best approach a particular situation and resolve a system error," adds Ren.

Results

Scaling the Data Science

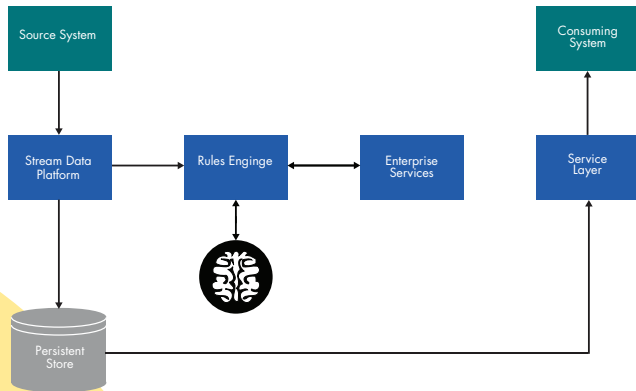
While the initial results of integrating machine learning into their systems have been promising, Comcast has set its sights on a greater vision – operationalizing big data analysis and building reliable predictive models that can scale to keep up with the company's volume of operations.

Among many issues that Comcast data scientists are working to overcome are extracting and integrating real-time production data that comes in different formats, using heavy computation to transform raw data into usable data sources, providing timely responses to a great number of prediction requests and continuously updating models with the latest data to keep predictions accurate.

Comcast is working with H2O to help overcome scalability challenges. "H2O enables us to operationalize more easily," continues Ren. "As we train a model, we can export it as a Java code or we can use H2O instances as a web service to build some intelligent apps based on our models."

"Interoperability with other toolsets is essential," concurs Leamon. "By using other tools to do some of the feature

Self-Healing System



"H2O.ai Machine Learning is one of the cooler things that we are working with in order to help us solve problems in a way that's really valuable for the organization."

– Drew Leamon, Director of Engineering Analysis at Comcast

engineering and have features prepared and ready to go into our models, we are able to plug them directly into Flow and other tools we are using."

At Comcast, Machine Learning and predictive analytics technologies are on the rise. "In the future, we will see more decisions leaning on this technology," concludes Leamon. "And not just executive decisions, we are seeing more and more intelligent systems and products that leverage this technology to do things like self healing, add better features to the product, predict what products will be trending in the future, and improve customer experience."

About H2O.ai

At H2O.ai we see a world where all software will incorporate AI, and we're focused on bringing AI to business through software. H2O.ai is the maker behind H2O, the leading open source machine learning platform for smarter applications and data products. H2O operationalizes data science by developing and deploying algorithms and models for R, Python and the Sparkling Water API for Spark. Some of H2O's mission critical applications include predictive maintenance, operational intelligence, security, fraud, auditing, churn, credit scoring, user based insurance, predicting sepsis, ICU monitoring and more in over 5,000 organizations. H2O is brewing a grassroots culture of data transformation in its customer communities. Customers include Capital One, Progressive Insurance, Zurich North America, Transamerica, Comcast, Nielsen Catalina Solutions, Neustar, Macy's, Walgreens, Kaiser Permanente and Aetna.

