Consensus simplifies data preparation for machine learning algorithm on AWS

CUSTOMER
Consensus

INDUSTRY
Retail / Software

USE CASE
AI / ML

ENVIRONMENT
AWS
Trifacta
Datarobot

Background

Consensus connects retailers with manufacturers and network operators using a single platform to enable what it calls multiplex selling: bundling technology and service purchases together, such as a smartphone and a data plan, or a smart television and a subscription to a video streaming service. Consensus enables retail stores nationwide to sell these types of connected devices and services through one unified online platform. However, a major risk for retailers selling expensive devices and services is fraudulent customer activity. To address this risk, Consensus rolled out fraud prevention as one of its core services. Through its automated machine learning-powered online engine, Consensus can alert its retailer clients to high-risk consumers before they purchase expensive devices.

To identify potential fraud, Consensus built an advanced machine learning model that leverages huge volumes of disparate data and undergoes routine updates. In order to be able to constantly refine its predictive models and quickly alert their retailer clients on potential fraud, Consensus sought out technologies that would allow it to prepare this data faster for use in its machine learning models.

Challenges

Preparing data for machine learning (commonly called "data wrangling") and creating machine learning models is a time-consuming process and poses a number of challenges for Consensus.

Business Challenges:
- The data preparation process required sophisticated knowledge of data science techniques, leaving the company's product and business intelligence teams unable to perform data preparation tasks on their own.
- Requests to their BI team (the single source of advanced SQL knowledge) for data preparation services took up valuable time and had to compete for resources that slowed down data exploration and innovation.

Technical Challenges:
- The painstaking process of re-engineering SQL scripts took Consensus up to six weeks (on average) to update its fraud detection machine learning model.
- Building machine learning models required tooling that was very text and command-line heavy, and required specialized knowledge that proved to be time-consuming.

Consensus realized they needed a better data preparation tool and rapid model prototyping and evaluation that did not require advanced data science knowledge from their product and BI teams.

> "With better and more accurate data, we knew we had the potential to save retailers a lot more money. Trifacta is really intuitive, and it's much easier to use than SQL/R solutions. Trifacta also complements our machine learning model solution very well."
>
> Harrison Lynch
> Senior Director of Product Development, Consensus Corporation

Solution

At first, Consensus attempted to work with several third-party data preparation solutions, but found them either expensive, unwieldy, or complicated to deploy. Finally, Harrison Lynch, Senior Director of Product Development for Consensus, discovered the DataRobot automated machine learning platform and free Trifacta Wrangler solution. Trifacta's user-friendly UI and DataRobot's ability to rapidly build and deploy machine learning models helped differentiate both technologies from the competition. Consensus then selected DataRobot and Trifacta Wrangler on AWS to obtain more computing power and connectivity in the cloud to wrangle more data, and quickly create machine learning models. "Retail fraud continues to rise and is one of the main causes of lost revenue—with each percentage of inventory lost to fraud, retailers lose roughly a million dollars in profit," Lynch says. "We are investing our product and processes in Trifacta and DataRobot to allow our business teams to pursue machine learning solutions as fast as they can think of them."

Benefits with Trifacta

By deploying Trifacta Wrangler and DataRobot on AWS, Consensus can now easily access data stored in both Amazon S3 and Amazon Redshift. In addition, Trifacta Wrangler leverages Amazon EMR (Elastic MapReduce) for data processing to take full advantage of the flexibility and scalability of AWS, allowing the company to leverage Trifacta's cloud data preparation and DataRobot's machine learning offerings with no large upfront infrastructure investment, while scaling as needed to meet the demands of their cloud workloads. Consensus can choose to deploy and manage its own AWS VPC (virtual private cloud) or select Trifacta or DataRobot managed cloud offerings. Consensus can also choose to deploy and manage their Trifacta deployment on their own cloud infrastructure through the AWS Marketplace.

Consensus was able to use Trifacta to wrangle large amounts of data stored in AWS and deliver more accurate results through DataRobot's machine learning algorithm faster than with the previous approach. The key benefits include:

Model Performance Improvement: 24% gain in True Positive detection, 55% decrease in False Positives, and 19% gain in overall financial performance
Fast Time to Deployment: reduce model implementation time from 3-4 weeks to less than 8 hours
Accelerated Model Training: from 2-3 days with at least 2 people to prepare data for each model, to training and validation in 2-3 hours by 1 person

With a joint solution of Trifacta and DataRobot in AWS, Consensus can now fuel high-quality data into Consensus's fraud detection models quickly, without having to go through time-consuming and costly manual data preparation process, resulting in faster time to more accurate result and improved overall customer experience.