Consensus Corporation, a subsidiary of Target, simplifies the complex process of selling connected devices. 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 adopted 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 data model that leverages huge volumes of disparate data and undergoes routine updates. In order to be able to constantly refine its predictive models and alert their retailer clients faster to potential fraud, Consensus sought out technologies that would allow it to prepare this data faster for use in its machine learning models.

The decision to use Trifacta and DataRobot

Preparing data for machine learning (commonly called “data wrangling”) and creating machine learning models is a time-consuming process. The painstaking process of re-engineering SQL scripts took Consensus up to six weeks (on average) to update its fraud detection machine learning model. In addition, 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 exploration and innovation. Furthermore, building machine learning models required tooling that was very text and command line

Trifacta & DataRobot: Automated data wrangling and machine learning for Amazon Web Services (AWS)

- Data preparation normally takes up to 80% of a data scientist’s time
- Trifacta- and DataRobot-hosted solutions on AWS drastically reduce the amount of time spent preparing and modeling data for machine learning
- Trifacta & DataRobot are AWS Machine Learning Competency Partners

CHALLENGE

Streamline time-intensive data preparation and machine learning tasks for use in predicting fraudulent activity at point-of-sale locations for a large national retailer.

SOLUTION

An automated data preparation and machine learning solution, built on AWS, that reduces the time needed to clean, structure, and ingest data and automatically build, train, and deploy machine learning models for predictive scoring.

RESULTS

Thanks to agile machine learning and the ability to prepare historical data quickly, Consensus achieved 24% improvement in fraud detection, 55% decrease in False Positives, and 19% gain in overall financial performance.

“Trifacta and DataRobot have helped reduce our feature engineering to model deployment from weeks to days, allowing us to respond to emerging fraud patterns quickly and accruing larger savings faster to our retailers.”

Harrison Lynch
Senior Director of Product Development,
Consensus Corporation
heavy, and required specialized knowledge and repetition of use that proved to be time-consuming. None of this was very user-friendly, and the result was lengthy and resource intensive.

Consensus identified the need for better data preparation and rapid model prototyping and evaluation that did not require advanced data science knowledge. At first, the company 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 Pro on AWS to get more power and more connectivity to wrangle even 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.”

### Trifacta, DataRobot, and AWS

Trifacta Wrangler Pro & DataRobot are AWS–deployed solutions that seamlessly access data stored on AWS, including Amazon S3 and Amazon Redshift. In addition, Trifacta Wrangler Pro leverages Amazon EMR (Elastic MapReduce) to process the data. The flexibility, scalability and superior economics of the cloud help companies try out Trifacta and DataRobot with no large upfront investment requirements, scaling as needed to meet the demands of their cloud workloads.

Companies can choose to deploy and manage the deployment on their own AWS VPC (virtual private cloud) or select Trifacta or DataRobot managed cloud offerings. Companies can also choose to deploy and manage their Trifacta deployment on their own cloud infrastructure through the AWS Marketplace.

### Trifacta Wrangler Pro and DataRobot:
Self-service data preparation and automated machine learning key features:

- **Data profiling**: Assess and resolve data quality issues up front to mitigate project risk
- **Transformation suggestions**: Structure and normalize data to fit standards to speed up project delivery
- **Advanced dataset manipulation**: Transform data of any complexity into a consistent view with no requirement for data engineers to manage the project
- **Interactivity and repeatability**: Real-time feedback for project execution efficiency and consistency
- **Build models**: Build machine learning models in one click using a massively parallel modeling engine that can scale to hundreds or even thousands of powerful servers
- **Validate**: Automatically search through millions of combinations of algorithms, data preprocessing steps, transformations, features, and tuning parameters to see which models perform best
- **Tune**: Automated model tuning with the ability to also do manual tuning to adjust machine learning algorithms for even better results
- **Deploy**: Deploy models for predictions with just a few mouse-clicks

“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
The benefits for Consensus

Consensus was able to use Trifacta to wrangle large amounts of structured historical data stored in Amazon S3 and more accurately deliver machine learning models in less time with DataRobot compared to traditional methods, translating in measurable benefits:

- **Model Performance Improvement**
  24% gain in True Positive detection
  55% decrease in False Positives, and
  19% gain in overall financial performance

- **Speed of Deployment** from three to four weeks of combined effort to less than eight hours to implement new models and respond to business changes

- **Accelerate Model Training** from two to three days and at least two people to prepare data for one model, to training and validation in only two to three hours by one person

Trifacta helped solve the problem of uploading the most accurate data into Consensus’s fraud detection models quickly, without the cost and potential inaccuracies associated with relying on manual data preparation or traditional languages such as SQL, R, or Python. DataRobot helped solve the problem of high numbers of false positive predictions that were hurting customer experience at the point of sale as well as detecting potential sources of fraud with higher accuracy.

ABOUT TRIFACTA AND DATAROBOT

Trifacta leverages decades of innovative research in human-computer interaction, scalable data management and machine learning to make the process of preparing data faster and more intuitive. Around the globe, tens of thousands of users at more than 10,000 companies, including leading brands like Deutsche Börse, Kaiser Permanente, New York Life and PepsiCo, are unlocking the potential of their data with Trifacta’s data wrangling solutions. For more information, visit www.trifacta.com.

DataRobot offers an automated machine learning platform that empowers users of all skill levels to develop and deploy machine learning and AI faster. Incorporating a library of hundreds of the most powerful open source machine learning algorithms, the DataRobot platform encapsulates every best practice, every safeguard, and every run-time environment to automate data science while maximizing transparency, accuracy, and collaboration, delivering AI applications at scale to leading brands including United Airlines, New York Life, LendingTree, and Tableau. Hundreds of other customers are using DataRobot to accelerate AI success. For more information, visit www.datarobot.com.

© 2019 Amazon Web Services, Inc., Trifacta, Inc., and DataRobot, Inc. or its affiliates. All rights reserved. DataRobot and the DataRobot logo are trademarks of DataRobot, Inc. Trifacta and the Trifacta logo are trademarks of Trifacta, Inc. All other marks are trademarks or registered trademarks of their respective holders.