

Data Science Case Study

Deploying Analytics for CPG Late Delivery Prediction

We designed and deployed an AI model to predict shipments at risk of late fees for one of the world's largest CPG businesses.



Industry

CPG



Use Case

Predicting Late
Product Shipments, ML
Deployment Challenges



Techniques

Analytics Adoption &
Scale, Model Evaluation,
Machine Learning



Outcome

Deployed model used
by operators

BACKGROUND

To maintain strategic relationships with retailers, consumer packaged goods (CPG) manufacturers must be able to reliably meet retail inventory demand. If supply chain delays result in the potential for late shipments, CPG manufacturers need advanced warning so that they can take action to avoid late deliveries.

A large player in the CPG sector came to Mosaic Data Science, an innovative data science consulting firm, with a late delivery forecasting model built from historical data. The company wanted Mosaic's expertise to transition from a statistical model – which, while useful in explaining the past, does not help predict the future – to a forward-looking predictive system that could be integrated with the company's existing IT infrastructure.

CPG DELIVERY PREDICTION

First, Mosaic's data science consultants discussed the current delivery planning process with executives at the company to understand the organizational needs for the model. It became clear that while the current model might be a good place to start, the model infrastructure needed to be modular enough that the forecasting model could be upgraded and easily replaced in the future.

In addition, Mosaic learned that the real-time data sources would be stored in [BigQuery and pre-processed in Google Cloud](#)¹. The client had deployed R to the cloud server for any necessary calculations.

Mosaic created a modular architecture for the predictive platform that can be run in two modes: a training mode that takes a model specification and a historical dataset as inputs and returns a trained model, and a scoring mode that takes a trained model and current data snapshot as inputs and returns a risk score (probability of late delivery) for each currently scheduled delivery. These output predictions could be viewed in a dashboard along with other information about in-progress or scheduled shipments so that high-risk shipments could be prioritized for action. The training mode architecture includes a validation step to provide estimated out-of-sample model performance statistics. Data inputs and model specifications are edited in standalone configuration files so that the customer can easily modify the data stream or model without risking changes to the overall framework. In addition, logs track the training and scoring processes so that there is a record of any errors introduced during the training or scoring procedures (e.g., unscored shipments due to missing values for some fields used in the model). Figure 1 shows the model architecture.



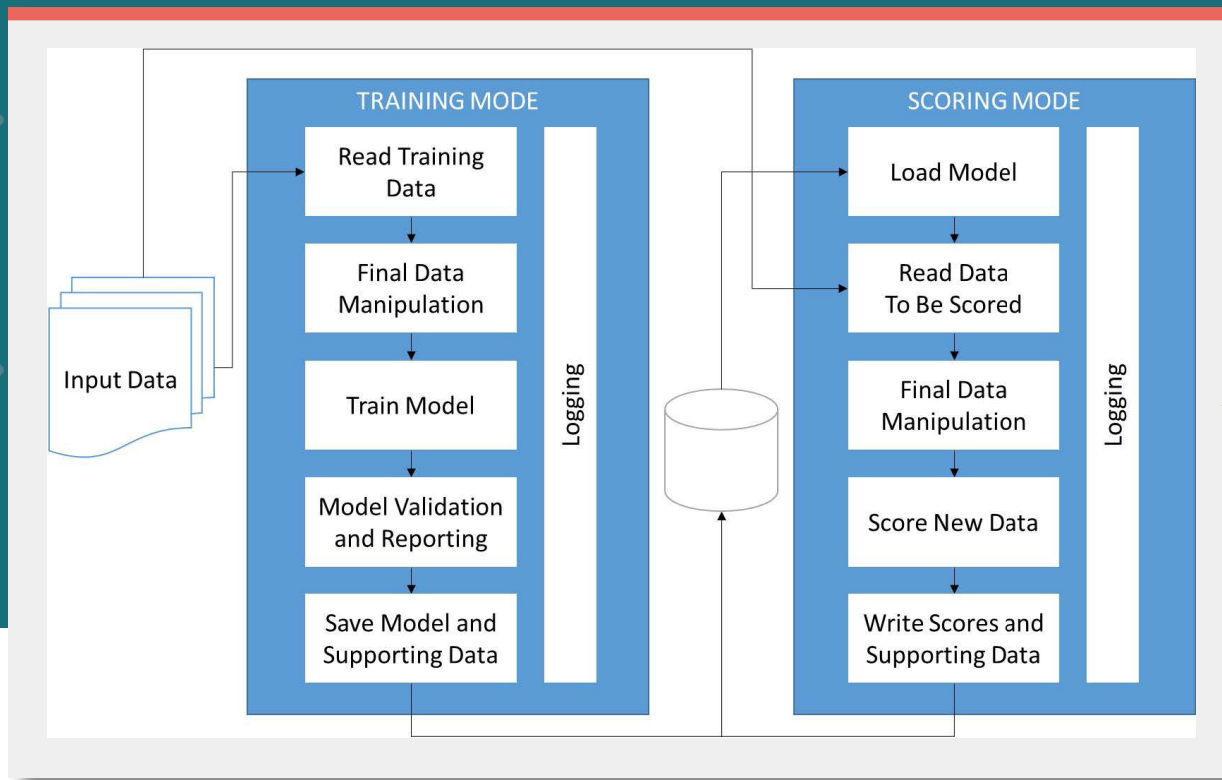


Figure 1. Delivery prediction model architecture to train model on historical data and create risk scores to highlight currently scheduled shipments at risk of late delivery.

RESULTS

Mosaic created a set of R scripts to execute the training and scoring modes with easily modifiable configuration files for each. In addition, Mosaic documented the code so that knowledge of the model implementation was not limited to the particular individuals at the company interacting with Mosaic’s team. The code has recently been linked to the CPG manufacturer’s BigQuery database to run in production. Mosaic Data Science remains engaged with the company, on-call every step of the way to advise on any implementation barriers.

Extracting insights from data is easy; making those insights actionable, particularly in an automated system, is difficult. These challenges are compounded when the analytics solutions are intended to integrate with existing enterprise technology. It is no surprise, then, that many companies come to Mosaic after spending significant effort developing state-of-the-art machine learning models that they are unable to use. By involving software development principles throughout the machine learning model development pipeline, Mosaic helps clients effectively navigate data integration hurdles and put successful analytics solutions in motion.



Mosaic's soups-to-nuts expertise—from model development through successful deployment—ensures that even clients with very complex IT systems can realize the benefits of predictive analytics in their businesses.

Endnotes

1. <https://cloud.google.com/architecture/data-preprocessing-for-ml-with-tf-transform-pt1>

