

What 'get the app' means for big data and AI

Originally published at <https://www.aiaa.net/artificial-intelligence/articles/what-get-the-app-means-for-big-data-and-ai>



Image: [Pexels](#)

New research undertaken at Imperial College University hints not only at the possibilities for machine learning to improve predictions about customer behavior, but at the huge benefit companies stand to gain by luring their customer base away from the web and into bespoke apps.

The [paper](#), *Customer Lifetime Value Prediction Using Embeddings*, provides a useful historical overview of machine learning efforts to identify high-value customers, before detailing a new research effort to apply deep neural network approaches to a database of 12.5 million customers—representing £1.15bn of annual revenue.

Very big data

The data in question belongs to global fashion retailer ASOS, a UK-based company which ships a range of 85,000 clothing items to 240 countries. ASOS is one of Europe's largest online-only outlets, operating in a sector where returns are an occupational hazard of business.

It's critical for a company with this business model to be able to identify high-value customers due to the potential cost of marketing or engaging excessively not only with 'zero value' clients (those who purchase but consistently return items, or have a stale purchase history, or who maintain a sporadic engagement with low-value items), but to 'negative value' customers, who return & refund so frequently as to represent an operating cost.

Read more: [Unlocking e-commerce success with artificial intelligence](#)

ASOS has historically used a Customer Lifetime Value (CLTV) model on principles which date back to the earliest methods of consumer analytics. These take into account customer demographics—in so far as they can be established—and a purchase and returns history.

However the researchers consider that web and app session logs represent "[by] far the largest and richest" of these data sources.

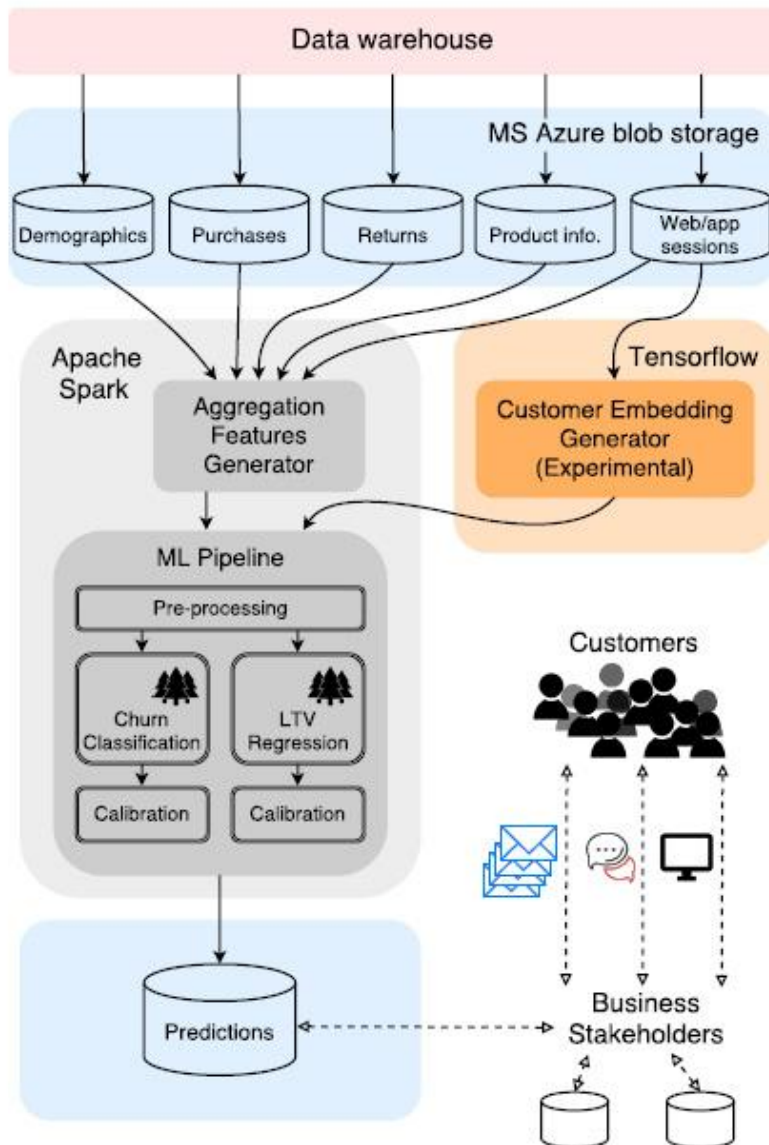
Feed it forward

The objective of CLTV research is to improve three key metrics—shopping frequency, average order size and churn rate. By identifying high-value customers, ASOS can undertake more targeted and economical retention strategies, with a focus on offsetting the weight of negative-value customers.

But most CLTV systems rely on resource-intensive handcrafted manipulation of the data, a logistics problem across the much broader field of general Deep Neural Network research.

The researchers, led by Benjamin Paul Chamberlain of ICL's Department of Computing, sought to discover how unsupervised learning could add a layer of utility to the model, and experimented with two approaches: the training of a 'feedforward' neural network on the manually-altered features within the data (supervised learning); and by adding unsupervised data derived from the study of customers' web and app browsing data.

As seen in the diagram, five types of information (demographics, purchases, returns, product info and web/app session data) were fed through Apache Spark, with Google's TensorFlow pre-filtering the more experimental session data, ultimately leading to predictions flowing through to stakeholders, potentially affecting the customer experience.



The model uses the previous twelve months of data, refreshed on a daily basis. This is necessary in order to account for known, seasonal fluctuations in customer spending; without considering this factor, the data being fed into the workflow could get skewed in either direction.

Considering such a large consumer database, the volume of data involved is breathtaking. Consequently potential new analysis models have to provide realistic estimates for resource usage. For the purposes of the experiment, the researchers found it necessary to establish averages and sampled data, generating templates which helped to group individual customers into broad categories.

Read more: [Lessons in augmented reality from the retail sector](#)

The paper finds that the deployment of web/app data through a neural network pipeline has enormous potential to improve CLTV predictions. The preliminary experiments produced an unusually close correlation between prediction of customer spending behavior and the real-world outcome.

However the researchers note that currently the costs of training and maintaining active neural networks towards these ends are likely to be prohibitive pending further advancements in DNN infrastructures and services.

The (mostly) unregulated app

Moving the customer from web-based services (such as a website) to an app represents the opening of an order of magnitude more data from the customer's interaction with the business. Depending on the platform, various permissions must be explicitly solicited from the customer either at the time of installation of the app, or later when it attempts to access facilities (contacts, camera, etc.) which have not been allowed yet.

However, the app can gather and disclose an extraordinary amount of data. Though this assumes that the user has given permission for the app to use bandwidth, in effect 'blocking' an app from the network either restricts functionality so far as to make the application useless; or else functionality is obstinately blocked because the developers have been ordered to get the data back to base.

This broad treatment of data sandboxing under Android and iOS is coming under [slowly-increasing scrutiny](#), however. Ridesharing behemoth Uber courted further controversy last year when it began to track the user's location five minutes after the finish of the ride, giving them no practical option to disable the functionality and claiming that it addressed personal safety issues. But the coming update iOS11 [will return](#) such 'out-of-scope' monitoring to the user's control.