



Rely on Automation for Scalability

Build an Effective Content Recommender System with Automated Data Cleansing and Scoring

Though it's common for online businesses to give recommendations for other services or products that consumers might be interested in based on their behavior, high-quality recommendations are extremely difficult to provide in practice. They can consume extraordinary amounts of internal resources, as they require heavy data manipulation, coding, and testing of different algorithms. So when a data team is tasked with creating a recommender system based off a variety of signals, the project can get complicated quickly.



Build an Effective Content Recommender System

About Our Customer

Our customer is a large national media organization that produces both news and public interest (cultural) pieces. They serve as a syndicator to a network of more than 500 stations across the United States. Additionally, their content is available directly from a variety of sources, including on the web and via their mobile application (which they develop and maintain in-house).

Industry

Media

Market

United States

Use Case

Recommendation System

Challenge

Build a Quality Recommendation System for Millions of Users Quickly

A large national media organization wanted to provide high-quality recommendations for users of their app. Their goal was to target consumers with content that they would actually be interested in based not only on what they previously consumed, but how exactly they interacted with topics in which they previously expressed interest. For example, if someone chose to listen to a report on Topic A but then fast forwarded through much of the piece (as opposed to actually listening to the piece in its entirety), the app should take that activity into account for future recommendations.

However, with a very small team and limited resources, the organization wanted to accomplish this in a scalable way. Not only would the system have to be mostly or entirely automated, but the team itself would have to be able to build the recommender easily in a way that would allow for quick tweaks and adjustments in the future.

Solution

Automate Process from Data Cleaning to Scoring

The team behind the national media organization's app leveraged Dataiku's collaborative data science and analytics platform to efficiently and easily string

together data cleaning operations and processes and then ultimately use that prepared data in the built-in supervised learning algorithms to build their recommender system.

Out-of-the box without having to write code, the team was able to seamlessly try out different models and choose the one that produced the best recommendations for the app according to their business goals. At the same time, when they did want to use code for data processing, Dataiku enabled them to use a combination of languages (mainly R and Scala), choosing the tool that would be easiest for the task at hand. And with all of the data processing required (sometimes more than 25 recipes for a single processing operation), if any part of their process failed, it was simple and efficient with Dataiku for the small team to tell exactly where the issue was and to fix it quickly.

Impact

200 Million Automated Predictions Every Evening

With automated data cleansing contributing to a data-driven recommender system, the media app automatically scores 200 million predictions every evening. Dataiku has made it feasible for the small team behind the media organization's content app to build and maintain this recommended system in a scalable way. This has allowed for an increase in productivity by 50%, therefore allowing the data science team to work on other projects in parallel.



Project Overview

Technology



Scala, Python, R

Time



6 months

Team



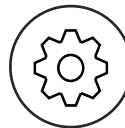
1 Data Scientist

Empower a small team to handle multiple data projects at scale



Quickly test machine learning models without writing code

Automate data cleaning and scoring processes for productivity gains



Identify and easily fix data processing issues in production

Project Overview: Building an Automated, Scalable Content Recommendation System

- > 200 million predictions automatically scored by the media app every evening.
- > 50% increase in data team productivity by using Dataiku to build, automate, and maintain the system.

GET STARTED