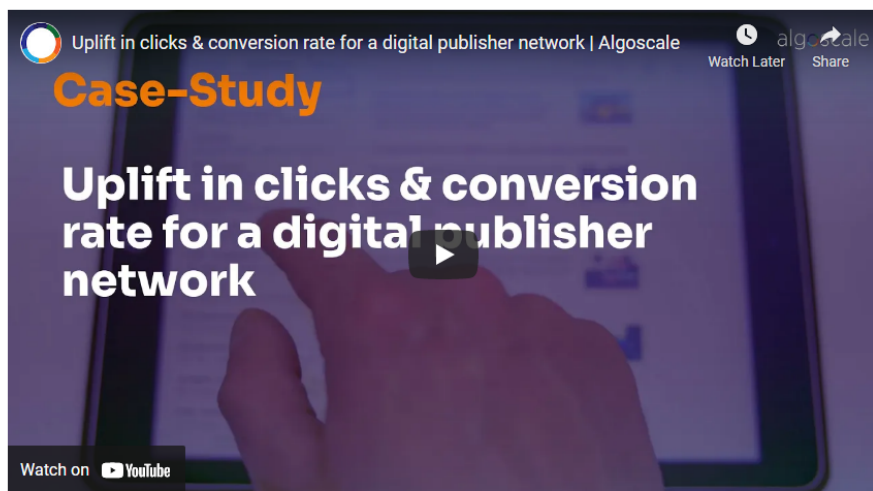


Built an intelligent and scalable ML model to increase CTR and subscribers for a media-tech company



Is there a time of the workday when you are more likely to read through travel websites? Or take time off your work to read an interesting article? Have you ever wondered why Facebook displays hotel ads just after you have searched for places to stay or even read a travel article? You might not know the answers, but whenever you are online, every click is being recorded and analyzed by media-tech firms to personalize your experience. Marketers use customer data platforms that provide them access to information to better understand customer behaviors, like when they engage with websites and what content they read.

Marketing has evolved from being "word of mouth" to digital. In digital marketing, click-through rate (CTR) is one of the most influential metrics to measure any campaign's effectiveness. But how do you do that?

Implementing marketing tools available in the market is a choice. But there are more than 6,000 marketing tools to choose from to understand user behavior. How do you make sense of such vast amounts of data?

To address a similar issue for a media-tech firm, Algoscale built an intelligent machine learning model to analyze site visitors' browsing habits, profile their personas and preference, and provide a 360-degree view to tailor customized content, thereby increasing the CTR and subscribers.

Client Overview:

The client is a media-tech company that helps online publishers convert website visitors to magazine subscribers. The company uses artificial intelligence and machine learning to integrate magazine articles into their websites using customizable, personalized widgets and generate user-preferred content.

Challenges and Requirements:

The media-tech company was looking for an intelligent solution to analyze thousands of online magazines and smartly segment the site visitors into the right categories based on their behavior. The audience segmentation would help them recommend widgets to the publishers to increase their click-through rate (CTR) and convert visitors to subscribers.

The firm also wanted a consolidated graphical report for an individual brand with details like nature, volume, and sources of web traffic, engagement trends, time on page, bounce rate, exit rate, top pages, CTR of widgets on the website, etc.

The Algoscale solution had to address the following challenges:

1. The firm did not have a separate storehouse for each of its clients. With all the client's data coming to a single data warehouse, there was a data swamp, making it difficult to segregate individual clients' data.
2. Some popular online magazines have massive traffic, which cannot be processed through classical Python and Pandas.
3. Backend data irregularities.
4. Validating our clustering and recommendation engine on real-life traffic was a challenge for the media-tech firm.

Solution:

Considering that each brand's website is unique, Algoscale built a universal ML model that can be optimized to analyze multiple websites independent of each other. To address the challenges highlighted by the client, Algoscale did the following:

1. Stored every online magazine publisher data in a separate table in a dedicated database to avoid the data swamp.
2. Implemented distributed computing of PySpark to process large volumes of traffic.
3. Implemented multiple new data preprocessing steps specific to the data to ensure data consistency.
4. Used Silhouette distance to validate the clustering model, which analyzed the separation distance between resulting clusters. Used uplift modeling to validate the recommendation engine, wherein data was divided into treatment and control to compute the uplift score.

The ML model had two components – user-behavior segmentation based on website engagement and widget recommendation based on user-behavior and four distinct modules to address the client requirement:

1. Data extraction and aggregation
2. Behavioral segmentation using a K-means clustering algorithm
3. Widget sequence prediction using Hidden Markovnikov Models (HMM) and Finite State Automata (FSA)
4. Deployment on AWS cloud server

Algoscale also built a pipeline using Python and Jupyter Notebook for generating reports for individual brands. The pipeline connects the client's database with the server that hosts Notebook and runs for a specific duration for a particular brand. Once the pipeline is executed completely, it presents a concise and insightful graphical report for that brand, which can be used to make informed data-driven decisions. The solution used the following tech stack:

Python, Pandas, PySpark, AWS Lambda, AWS EC2, AWS S3 buckets, Dynamo DB, Scikit-learn, MySQL, RDS, Seaborn, and Matplotlib.

Benefits:

The solution provided the following benefits to the media-tech firm:

- Reduced customer acquisition cost across multiple publishers.
- Increased CTR substantially across various publishers.
- Customized content recommendation based on website engagement data.