

AI in Media & Entertainment



A WHITE PAPER BY DATAIKU

www.dataiku.com

TABLE OF CONTENTS

- 2** • Introduction
- 3** • High-value AI use cases in media and entertainment
- 8** • AI in a nutshell for M&E industries
- 12** • Feature: Success stories from leading M&E companies
 - 12** • Ubisoft
 - 19** • DAZN
 - 21** • Betcltic
 - 24** • Infopro Digital
- 26** • Main AI challenges in media and entertainment
- 29** • Upcoming trends in AI and machine learning for M&E
- 33** • Conclusion



INTRODUCTION

Data has always played an important role in the media and entertainment industry. TV and radio programming lives and dies based on ratings generated from viewers and listeners. Long before newspapers were monitoring clicks and social media shares, they were carefully tracking subscriptions and analyzing the demographics of their readers.

As content consumption behaviors are becoming increasingly complex and evolving more rapidly than ever, media and entertainment companies are facing increasingly competitive and uncertain markets, which are driving the need to reduce operating costs and simultaneously generate more revenue from delivering content. Data science, machine learning (ML), and AI represent huge opportunities for companies who are prepared to invest in long-term viable solutions.

Companies, in turn, are tailoring their offerings and business models to revolve around personal preferences, leveraging data and usage patterns to pitch their products not at audiences of billions, but at billions of individuals.

“The central theme of this growing world of media is that it’s personal and increasingly digital. And it is one that is constructed by the individual for his or her own enjoyment and gratification, and delivered through personal devices.”

*“Getting personal: Putting the me in entertainment and media”,
Global Entertainment & Media Outlook 2019–2023, PwC*

The internet has magnified the role of data in media. Every player in the industry, from a solo blogger to multinational media conglomerates, has easy access to an unprecedented amount of information about existing and prospective customers. Furthermore, they have an unprecedented ability to use data to target certain customers. The largest platforms cater to and anticipate users’ personal needs with AI-enabled, algorithm-driven playlists and recommendations. In response, marketers are racing to reach individuals at the point of consumption.

Just look at the way that companies like Google, Amazon, and Netflix affected the media and entertainment space with their data-driven business models. Using advanced data science and AI allows them to understand and create value for customers by presenting content that they might like and that might be relevant to them. But this change doesn’t happen overnight; full potential comes through gradual evolution into a truly data-powered organization.

This paper will delve into...

- an in-depth exploration of the highest value AI and data science use cases in media and entertainment;
- a breakdown of the key AI opportunities and challenges by industry;
- a selection of AI success stories from prominent M&E companies;
- an overview of the main challenges and upcoming trends for AI in media and entertainment.



I. HIGH-VALUE USE CASES IN THE MEDIA AND ENTERTAINMENT INDUSTRY

SMART RECOMMENDATION ENGINES

Recommendation engines have been widely used in the media industry to predict what kind of information or content customers would be interested in. Companies can combine structured and unstructured data and machine learning methods to match people and content, thus improving the relevance of content recommendations and efficiency of content distribution.

With leading tech media players such as TikTok and Netflix venturing more and more into AI-based interactive and smart content, we're likely to see a shift from simpler content recommendation systems to an entire AI-driven personalized content experience.

FURTHER READING

How To: Build a (Quality) Recommendation Engine



HYPER-TARGETED ADVERTISING

The possibility of combining data from different sources in one place can allow companies to look at their customers as a whole and deliver unique, hyper-targeted offers. In TV and advertising, this is evoked in the concept of addressability: the ability to interact with consumers based on what their specific choices reveal about their interests and preferences. Hence, thanks to AI and ML, media and entertainment companies can predict churn rates more accurately, place advertising at the right time and in the right place, and have more appropriate, personalized offers to increase conversion.

For instance, some streaming platforms and leading film studios are already experimenting with ML-based personalization of movie trailers that emphasizes specific elements that they know a given target audience would like, delivered on the platform that they use most frequently. Last November, 20th Century Fox explained how it used AI to detect objects and scenes within a trailer and then predict which "micro-segment" of an audience would find the film most appealing.



REAL-TIME PREDICTIVE MODELLING FOR ANTICIPATING DEMAND AND SEGMENTATION

In the constantly evolving media and entertainment sector, looking back at consumers' past activity often isn't a good indication of what they will do next. Instead, real-time prediction based on current trends and behaviors from all data sources is key.

Predictive modelling in particular will aid media and entertainment companies not just by allowing them to react to consumers in real time, but also to anticipate their behavior, influencing long-term investments, for instance, what kinds of movies in which consumer micro-segments will be popular two years from now. In addition, companies can make predictions about which customers are more likely to view a given type of content, and what device they will be using when viewing it.

PROGRAMMATIC AD BUYING

Traditionally, companies purchase advertisement slots for TV programs long before the shows air. While those ad buys are usually based on an analysis of audience data (age/gender/geography etc), they do not account for the high level of fluidity in viewership. The ad buying process itself is cumbersome. This is how Google Double Click described the status quo in a [recent white paper](#):

While this traditional TV buying and selling model has worked for decades, it's not without its inefficiencies. Chief among them is the fact that many of the interactions that occur are manual: requests for proposals, insertion orders, and ad trafficking, not to mention the endless emails, spreadsheets, and even faxes.

Hence the arrival of "programmatic ad buying," which leverages real-time data analysis and automation to purchase ads across a wide variety of media platforms: broadcast TV, cable, satellite, over-the-top services like Hulu and Netflix, and online video services like YouTube. This new method of ad-buying will involve a system that can constantly monitor audience dynamics across multiple channels and respond purchase ad space as soon as it becomes available.

PERSONALIZED PROGRAMMING

There is perhaps no more prominent example of the impact of AI in media than over-the-top services such as Netflix, Hulu and Amazon Prime. These platforms have not only transformed the way people consume film and television through on-demand programming with significantly fewer advertisements, but they have transformed the way that programs are marketed. Every single one of their customers are providing a steady stream of data that paints a sophisticated portrait of their preferences and viewing habits.



Every time somebody logs into their Netflix account they are presented with a personalized set of programs based on their viewing history. The platform's algorithms use a number of different data points to predict what a person will like. Not only will it identify patterns in your viewing history (that you have sought out multiple sports documentaries or French-language films, for instance), but it will recommend programs that that others with similar viewing histories have watched. The algorithm is constantly improving based on the data fed to it by millions of Netflix users.

Netflix has similarly employed AI to optimize the way that movies and shows are presented to users. Netflix tests different thumbnail displays for programs to understand which ones are most likely to [attract viewers](#). Again, pairing this with other viewership data may lead the platform to conclude that certain designs work better on certain viewers.

AUTOMATING HR, LEGAL AND ADMINISTRATIVE TASKS

Media and entertainment companies could also apply AI and automated machine learning in improving the efficiency and reducing the time and cost in a number of HR, legal and administrative tasks.

Residual payment forecasting refers to the way talent is compensated based on content distribution across various channels. For movie studios, determining and forecasting residual compensation for each contributor can be a huge sinkhole of time and effort, but the good news is that large parts of this process could be automated. A properly trained AI or autoML algorithm can do a lot of the grunt work in such research, by seeking out key clauses and provisions in contracts related to residual compensation, as well as by automatically tracking and analyzing how different cuts of the same movie will affect each actor's residual payment.

Ensuring compliance with copyright standards. Another tedious task for TV studios and distributors in particular is verifying whether they still have the rights to the characters, music, script, props, etc. for every rerun of a given show or emission. AI has the potential to free expensive legal talent from reading thousands of documents by seeking out every instance of relevant keywords in those reams of data.

AI-Based casting: Some companies are starting to experiment with using AI for casting optimization purposes. A Los Angeles-based startup licenses historical data about movie performances over the years, then cross-references it with information about films' themes and key talent, using machine learning to tease out hidden patterns in the data. Its software lets film creators play fantasy football with their movie, inputting a cast, then swapping one actor for another to see how this affects a film's projected box office.

FURTHER READING

AI for Human Resources



AUTOMATED “ROBOT REPORTING” TO FREE UP JOURNALISTS’ TIME FOR MEANINGFUL WORK

If you thought journalism was a uniquely human profession, think again. An increasing percentage of online news content was not written by people but by robots with advanced natural language processing capabilities.

A third of the articles on Bloomberg, the popular business news site, are at least partially written by Cyborg, an AI-based program which can dissect a financial report the moment it appears and spit out an immediate news story that includes the most pertinent facts and figures. The best part is that, unlike business reporters, who usually find that sort of repetitive task unfulfilling and boring, Cyborg does it without complaint.

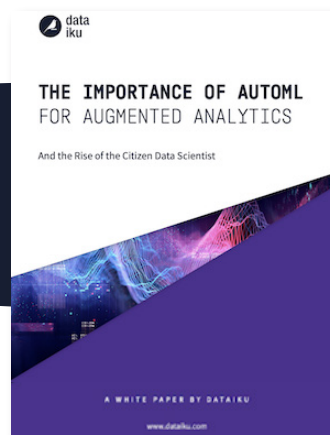
Another major news outlet that has ventured into robot reporting is the Associated Press, which has used it for automated financial news coverage and to produce articles on minor league baseball.

The Associated Press (AP) estimated that robot reporting has freed up 20 percent of reporters’ time spent covering corporate earnings and that AI is also moving the needle on accuracy. “In the case of automated financial news coverage by AP, the error rate in the copy decreased even as the volume of the output increased more than tenfold,” said Francesco Marconi, AP’s strategy manager and AI co-lead.

Since 2016, the Washington Post has used their homegrown robot reporter Heliograf to comb through data and identify trends in business, politics, weather or any other topic that a reporter may be interested in covering. In essence, Heliograph serves as a highly effective researcher in service of journalists¹.

Media outlets using AI say it’s meant to enable journalists to do more high-value work, not take their jobs. There is still substantial news content where robots can assist, but not replace humans. People don’t read newspapers and magazines just because they want information. They also seek analysis and insight that leverages human reasoning and emotion that a computer cannot easily replicate.

FURTHER READING Automation and the AI Skill Gap



VIDEO GAME DEVELOPMENT

There is perhaps no more vivid demonstration of the power of AI than in video games. Indeed, what better displays the potential of artificial intelligence than artificial humans engaging in human activity, whether that's playing basketball or waging war?

It is AI that explains why video games have become steadily more sophisticated in recent years, with computer-generated opponents who appear to think and act like humans, reacting to and anticipating your actions, whether it's a football player avoiding your tackle or an enemy soldier ducking your fire.

One common AI technique that has been used in video games since the 1990's is Finite State Machine, an algorithm that programs non-player characters with a reaction for every possible situation created by a player. FSM was included even in relatively simplistic, two-dimensional games. The algorithms that guide non-player characters in today's highly-sophisticated 3D games, however, take into account exponentially more scenarios.

The next step up in sophistication is the Monte Carlo Search Tree. MCST was the foundation of Deep Blue, the IBM chess machine that shocked the world by beating chess master Gary Kasparov in 1997. MCST enabled Deep Blue to consider every single move it could make and then consider every single move its opponent would make in response. Each potential move and potential response creates a different branch and then a wide variety of other branches as the algorithm considers the countless sequences of events that could follow.





FEATURE

II. AI IN A NUTSHELL FOR M&E INDUSTRIES

MUSIC

For a long time, the music industry had been in need of disruption, and the emergence of data and AI technology into the sphere seemed to fulfill this. The shift to a data-driven business model in the face of music subscription services represents one of the largest industry transformations that music has witnessed in many years, completely upending the revenue model of music companies.

The growth of streaming services such as Spotify and Apple Music has strongly accelerated that change. Most executives now understand the value in analyzing user behavior and adapting strategies based on those metrics.

Core use cases:

- Recommendation engines
- Trend and sales forecasting
- Customer segmentation
- Monetization and payment forecasting

Check out our bonus feature interview with a data scientist at [Spotify](#) about the power of data storytelling in the music industry. You can find it on page 31.





FILM & TV

We're only at the beginning of the technology-based radical transformation of the media and entertainment industry. Compared with the music industry, film and TV companies have not been through their big transition yet. Some of it is already happening with streaming and subscription services, but most film studios still operate mainly in the traditional model.

Streaming services are at the forefront of the AI and data science revolution. Production companies, including Amazon, Hulu, and Netflix, analyze patterns in big data to determine the types of content they create and make personalized viewing recommendations. In this way, data science can aid the art of producing and marketing entertainment at levels never before seen.

Core use cases:

- Content recommendation systems
- Churn prediction
- Hyper-targeted marketing and advertising
- Sales and residual payment forecasting
- Social sentiment analysis and trends forecasting

*See how **DAZN**, a company providing sports subscription services, successfully scaled a small data team by leveraging the power of machine learning, on page 19.*





VIDEO GAMES

Data has always played an important role in media and entertainment. This is especially true for the gaming industry, where generating sophisticated customer insights is essential to understand the players' perceptions of the games and develop highly personalized and effective product strategies.

Millions of people are playing video games at any given moment all over the world. Gamers are constantly interacting with the games themselves, with game creators and influencers, and with each other, all of which generate a significant amount of valuable customer and social data. Being able to effectively transform all of this data into actionable business insights is key for game developers.

Core use cases:

- Game development
- Game design, visual effects and graphics
- Object identification
- Personalized marketing
- Social and customer analysis

*Check out page 12 to see how **Ubisoft**, one of the world's top video game companies successfully nurtured and upscaled their data team, thus empowering them to understand their customers better and develop meaningful and personal solutions.*

NEWS

The growing concerns over the long-term profitability of both print and digital news media has left news organizations scrambling for answers. Now, more than ever before, publishers need to reinvent their models in ways that would allow them to reach their target audiences in smarter, more efficient ways. What's more, they need sustainable solutions for monetizing their content.

Nowadays there are more customer data sources and analytics tools available for online publishers than ever before in the media industry. These unique sources provide companies with massive amounts of data, allowing them to have accurate and comprehensive insights on their customer base, its demographics, behavior and preferences.

Core use cases:

- Improving ROI and revenue model
- Sales forecasting
- Hyper-targeted advertising
- Robot reporting





SPORTS ENTERTAINMENT AND GAMBLING

The global online gambling market is anticipated to be valued at more than \$94 billion in 2024, and increasing use of more advanced data science, machine learning, and AI techniques to improve the online gambling experience will most certainly play a role in this market growth.

In fact, one of the industry's biggest challenges could be what to do with all of the masses of data at their disposal and whether they can turn all of that data into actual business value. Unlike traditional sales use cases, data science in online gambling means not only personalization and anticipation of individual behavior, but also taking into account the uncertainty surrounding any action the user takes.

Core use cases:

- Fraud detection
- Lifetime customer value prediction
- Ad-hoc analysis on data from A/B tests

*See how **Betclix**, the leader in online gambling on the French market, is rising to these challenges, on page 21 .*

ADVERTISING

Media advertisers have long relied on audience data to determine when and where to place their campaigns. Recent innovations in AI and data science, however, has dramatically increased advertisers' understanding of their audience. While some parts of the media and advertising industries have been slow to embrace change, it's only a matter of time before innovations fundamentally transform the way ads are bought and sold.

Core use cases:

- Hyper-targeted advertising
- Programmatic ad buying
- Customer segmentation

*Go to page 24 to see how global media company **Infopro Digital** transformed its organization around data to deliver one of the most advanced business-to-business advertising targeting engines on the market.*



FEATURE

III. SUCCESS STORIES FROM LEADING M&E COMPANIES

In this section, you can learn about how leading media and entertainment companies are leveraging AI, machine learning, and data science tools (by using Dataiku) in order to increase their competitiveness and extract real business value from their data initiatives.



UBISOFT

COMPANY INFORMATION

- The third largest video game company in the world
- +40 studios and offices in +95 countries around the world
- +17000 collaborators
- Develops games for franchises such as Assassin's Creed, Far Cry, Prince of Persia, Just Dance and Rabbids.

TOOL & TECH STACK



VINCENT NICOLAS

Data Scientist Manager

Primarily responsible for: making the link between the IT department and the technical needs of the CMK (Consumer and Market Knowledge) department; data visualization expert and constant technological watch.

Main goals: break down data silos, facilitate the access of the CMK team to all of the available Ubisoft data (internal and external). Empower the CMK team to generate powerful insights to improve the overall product quality and respond to player needs.

THE DATA TEAM

Ubisoft's Consumer and Market Knowledge (CMK) department is in charge of market research. Its primary function is providing player and market insights across the company in order to improve the overall quality of games and effectively respond to the emerging player demands.

ORGANIZATION

The Consumer and Market Knowledge department is comprised of around 50 collaborators, among which 10 data scientists and 30 analysts.

The CMK department works with 3 different entities:

- **Studios:** provide them with insights for game development
- **Top management:** help them navigate the constantly changing trends in the gaming space and help define their strategic vision
- **Business:** help each game find the right audience





STRUCTURE

DATA SCIENCE

Split in 2 teams: Operational and Research.
Primarily responsible for:

- Generating insights for the product development team
- Competition analysis and prediction
- Sales prediction
- Analyzing social data from forums, online gaming communities
- Motivational models



MARKET KNOWLEDGE AND TRENDS

The Market Knowledge team analyzes video game market data, and business models evolution.

The Trends subdivision of the CMK is in charge of anticipating, analysing and forecasting the emerging expectations of target players, as well as more broadly, the new generation of players.

CMK ANALYSTS

The CMK analysts are primarily responsible for analyzing typologies in user behavior in order to help production and business teams through:

- Telemetry data (collected via tags): how do gamers play a particular game, and why, what are their motivations and emotional drivers.
- Declarative data (regular and punctual surveys)
- Crossing qualitative, quantitative and tracking data.





THE CHALLENGES

- Rapidly changing profiles and expectations;
- Very different user profiles and patterns: many different styles and ways of playing one single game, different motivations (individual missions, overall story development, customization, social aspect...).
- Data viability and complexity: sometimes, there is contradiction between the behavioral/telemetry data and the declarative data
- Lack of standardization in methods and tools used by various profiles
- Need to democratize access to data and facilitate self-service analytics.

THE MISSION

The CMK department's mission is fully aligned with Ubisoft's : enrich the lives of video gamers. But in order to do that, it's crucial to know them well, as well as the market and the environment in which they evolve.

Main goals

- Understand all the fine undertones of gamers' behavior and needs, and generate insights for product and marketing;
- Leverage the potential of telemetry data not only for product optimization, but also for marketing and business purposes
- Elevate analysts' specialized knowledge in consumer profiles by empowering them to use data science tools

THE SOLUTION

Embarking the CMK analysts into the world of data wasn't the easiest mission, but with a consistent and comprehensive step-by-step approach, the CMK department was able to successfully scale a diverse collaboration-driven data team and implement a powerful data-driven market research strategy.



THE ODYSSEY OF A DATA TEAM

2018

Embarking on the data journey

Equipping the team with the right tools

After conducting a comprehensive study and identifying the main pain points of the CMK, Ubisoft realized that rather than trying to train analysts on data science tools, they need a transformation in tools and practices that would foster collaboration and leverage analysts' skills and knowledge in a more comprehensive way. Soon after integrating an initial point-and-click metabase, the CMK department started to implement Dataiku's solution.

Embarking the crew for the (DSS) enterprise

To better tackle the data and research needs of the company, as well as those of the consumers, the CMK split their data science team in two subdivisions - operational and research. An additional data scientist was hired, and after only a few weeks, the first projects started rolling out.

Fostering a data-driven culture

A series of trainings was conducted, which leveraged analysts specific areas of expertise and allowed them to build their first end-to-end projects.

Additionally, the department started to foster and nurture a positive data culture change across teams, by partnering with Dataiku on internal events and data competitions.

2017

A need emerges:

First initiative to give analysts access to exploring telemetric data and be able to use cross analysis between telemetric and declarative data.

Understanding analysts' needs

An internal study was launched to understand the analyst teams' needs, skill sets and motivations in terms of working with data and advanced analytics. The study showed that there is a real desire among analysts to get into data collection and exploration, as well as understand the basics of machine learning.

First trainings

A number of analysts started getting trained on Python, R and SQL, but it was taking too much time away from their primary tasks and it interferes with the group team dynamic. There was an apparent need for an easier, more effective solution.

2019

Sailing to new horizons

Transitioning from collaboration to co-elaboration

Each department has its own collaborative space on Dataiku and develops its own projects. The analysts are not only able to easily access the data sources of the data science team, but are also able to develop their own data projects.

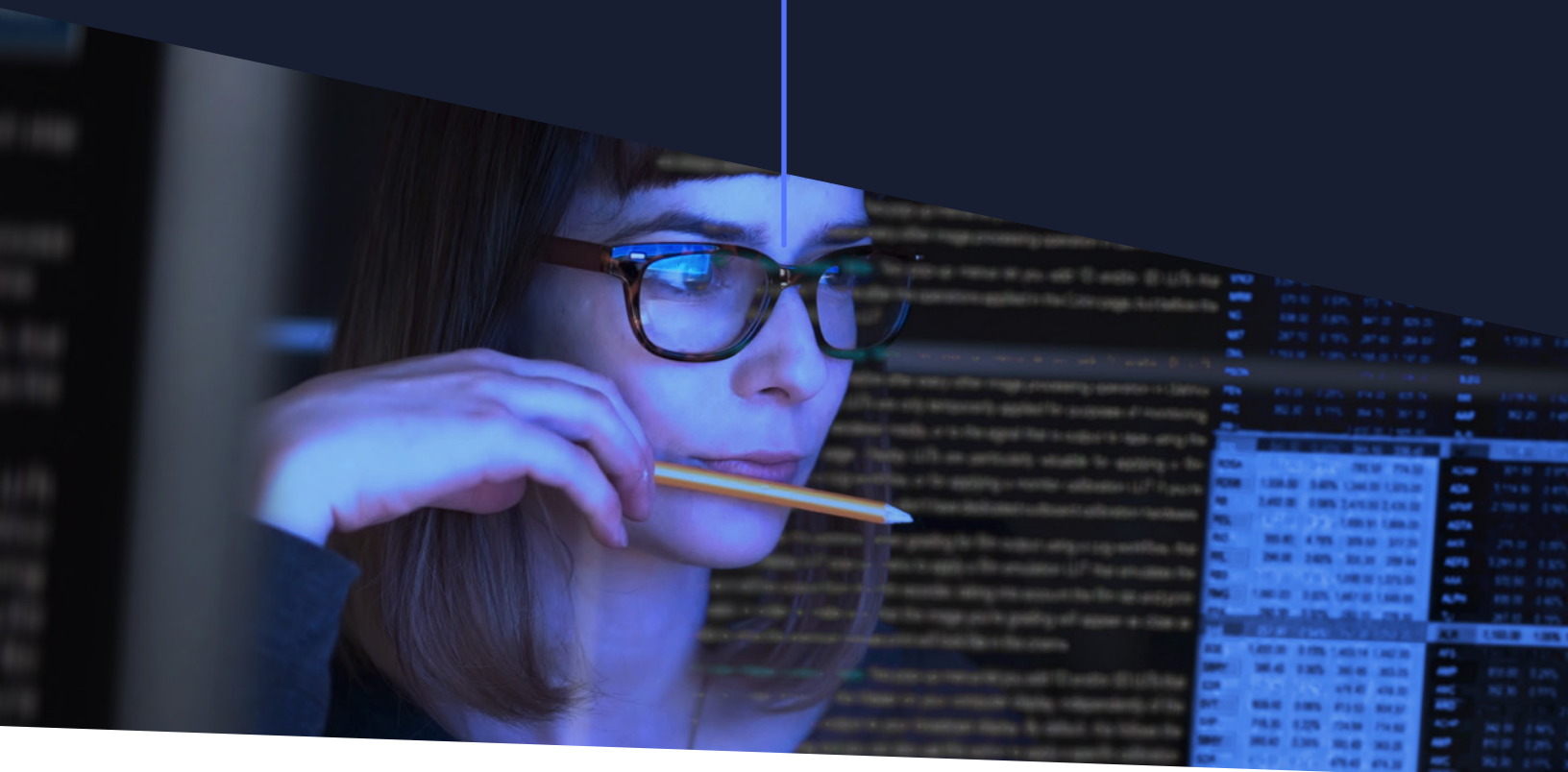
Upscaling data skills.

In October 2019, another 25 collaborators are trained on Dataiku's design node.

An analyst successfully transitions towards the data science team thanks to the skills and knowledge developed using Dataiku.

“Due to Dataiku’s easy to use point-and-click interface which empowers analysts to easily connect to data and collaborate on data projects, a real curiosity and data-driven mindset emerged across the whole department.”

Vincent Nicolas
Data Scientist Manager at Ubisoft



THE RESULTS

Thanks to Dataiku, Ubisoft was able to:

- Upscale analytics efforts, give access to data and analytics to non-coders
- Facilitate collaboration between business analysts and data scientists, as well as between the data team and lines of business
- Build sales and competitor prediction models
- Do cross-comparative analysis of player behavior and market knowledge to predict consumer trends and generate actionable business insights

Dataiku's strengths

- Facilitates and encourages collaboration between the data science and the analyst teams
- Gives easy access to various data sources for the whole department (Teradata, Hadoop cluster)

"Dataiku is the ultimate collaborative data science tool. It gives analysts autonomy, while also offering data scientists enough freedom to explore."

Vincent Nicolas
Data Scientist Manager at Ubisoft

Dataiku has been an accelerator for the realization of the projects and the transformation of Ubisoft's CMK team. This transformation implies an evolution of the organization, of the collaboration between the teams, and a positive cultural shift around data that is felt across the company.

"Dataiku listens to us and understands our needs the way we like to think that we listen and try to understand our players."

Vincent Nicolas
Data Scientist Manager at Ubisoft



SCALING A SMALL DATA TEAM WITH THE POWER OF MACHINE LEARNING

COMPANY INFORMATION

DAZN is a subscription service owned by Perform Group dedicated to live and on-demand streaming of worldwide sporting events. It offers access to more than 8,000 sporting events a year across a wide range of devices to customers in Austria, Germany, Japan, Switzerland, and Canada, with more markets coming soon. In an effort to continue to grow their business in existing and new markets, DAZN wanted a fast, low-maintenance way to enable their small data team to run predictive analytics and machine learning projects at scale.

THE CHALLENGE

In an effort to continue to grow their business in existing and new markets, DAZN wanted a fast, low-maintenance way to enable their small data team to run predictive analytics and machine learning projects at scale.

In addition, they wanted to find a way to allow data analysts who were not necessarily technical or experienced in machine learning to be able to contribute in meaningful ways to impactful data projects. Ultimately, they wanted to support an underlying data culture with advanced analytics and machine learning at the heart of the business



THE SOLUTION

DAZN knew that in order to accomplish their goals quickly, they would need technologies that were simple and in the cloud. They turned to Amazon Web Services (AWS) and Dataiku in combination for their simplicity in setup, connection, integration, and usability, and they got up and running in under one hour.

With AWS and Dataiku, the small data team built and now manages more than 30 models in parallel, all without needing to do any coding so that the processes are completely accessible to non-technical team members. They use these models as the basis for a variety of critical processes throughout all areas of the business, namely:

- Content attribution to determine what fixtures are driving sales, enabling contextual information on key fixtures in each market.
- Advanced customer segmentation to identify user behaviors, particularly regarding content and devices on which customers use the product.
- Propensity modeling to identify customers that are likely to churn, enabling improved customer targeting for retention activities.
- Survival analysis to understand customer stickiness, enabling calculation of expected revenues to understand customer return on investment.
- Natural language processing on social networks for market research.

THE IMPACT

AWS and Dataiku have noticeably shifted the data culture at DAZN and have brought innovations in advanced analytics and machine learning into the spotlight throughout the company. Thanks to Dataiku's ease, simplicity, and huge efficiency gains, DAZN has hired two data analysts who have already gotten up to speed and are doing as much work as five analysts in the pre-Dataiku team.

Overall, the biggest impact has been empowering a non-technical team to create more models than ever before and get them into the production environment quickly to bring real ROI to the business. DAZN plans to continue to grow the team to three data scientists and 6-10 analysts to exponentially increase the number of machine learning models in production.

FURTHER READING

How to Nurture a Productive Data Team



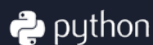
PUTTING DATA SCIENCE AT THE CENTER OF ONLINE GAMBLING



COMPANY INFORMATION

- Founded in 2005 and currently the largest betting company in France
- Specialized in sports betting, horse race betting, online casino, and online poker

TOOL & TECH STACK



BERTRAND WAROT

Data Science Manager

How Betclic Uses Dataiku

- Connection to data and data preparation
- Collaboration between business analysts and data experts (data scientists, engineers, etc.)
- Testing and iterating on machine learning models quickly (no code required)

In just 1.5 years, Betclic has created more than 130 Dataiku projects, including those powering dashboards, ad-hoc analysis, and more.





USE CASES

The core of every data project at Betcllic starts from a business need, which might come from any number of departments - marketing, finance, fraud, or customer service. From there, data teams prioritize those requests from the business and work with them to flesh out their needs (which may be different from their initial questions or requests for deliverables).

“Our goal is to respond to the needs of our internal clients more so than their asks.”

Bertrand Warot, Data Science Manager at Betcllic

Some examples of use cases the data team at Betcllic has worked on include:

- Anticipating new customers' future value
- Ad-hoc analysis on data from A/B tests
- Money laundering detection
- Linking related fraudulent accounts

DEVELOPMENT OF A DATA DEPARTMENT

Yet tackling a wide range of use cases and attempting to put data science at the center of all operations didn't come naturally - it required some fundamental organizational change. Betcllic's CEO created the data department in 2017 in an effort to do just that - make data the core of all Betcllic projects.

Since then, the team has grown from just three people in the R&D department to a proper data department of more than 30 people who are responsible not just for one-off projects, but the overall management and incorporation of data into processes. The department structured around three areas: data engineering, analytics, and data science, all of which sit under a Chief Data Officer (CDO).

“In the context of rapid expansion of the data department, Dataiku supported us in upskilling analysts while also enabling us to ensure the same level of productivity with the already-established team.”

Bertrand Warot, Data Science Manager at Betcllic





PROCESSES AND KEYS TO SUCCESS

Along with the organizational change required to upskill people and staff a data department, Betcltic was able to better tackle some of its use cases by making some process changes as well.

BEFORE

In the early days of data science efforts, access to data was controlled through written request to the database administrators, after which they would deliver datasets as CSVs (which may or may not correspond to expectations or the data needed for the project).

Each team member worked independently with his or her own processes, organization, and versioning, and files were stored on local machines.

AFTER

To scale their team, the data department needed a way to eliminate maintenance and communication issues caused by their previous setup. Betcltic turned to Dataiku to help centralize all data efforts plus offer a solution to sharing of projects, versioning, archiving, and automation. As the team continues to grow and junior analysts come with new skills, Dataiku enables Betcltic to centralize all of that work and train staff to handle more (and more complex) parts of the data-to-insights process.

"Thanks to Dataiku, we all march to the beat of the same drum: we are able to manage projects from A to Z, which makes it possible to overcome all the problems of automation and scripting."

Bertrand Warot, Data Science Manager at Betcltic



HARNESSING DATA TO SERVE MORE RELEVANT ADS

COMPANY INFORMATION

Infopro Digital is a cross-media company founded in 2001 that focuses on media outlets, events, lead databases, and lead generation. The company has about 2,800 employees with a turnover of around €360 million. They manage 50 media outlets as well as 300 events annually. Infopro Digital is a worldwide company, but they are especially prominent in the United Kingdom, China, Spain, France, and the United States.

THE CHALLENGE: PROVIDE BETTER TARGETING WITHOUT INTRODUCING DEPENDENCE ON IT

Infopro Digital owns a wide range of leading business-to-business digital media outlets, and they sell advertising space on these sites. To differentiate itself from the competition, Infopro Digital wanted to offer more advanced targeting options to its advertising customers. For example, instead of doing basic category targeting (like showing aerospace ads in the aerospace section of a website), they wanted to leverage the user's navigation path and behavior to more accurately target those who may be interested in a particular ad, wherever they are. This advanced targeting requires experienced technical teams to handle a vast data lake. But despite the highly technical environment, Infopro Digital's marketing teams needed to be able to handle the queries and most of the day-to-day work themselves without the help of IT every time. The marketing teams had some prior knowledge of processing data using Microsoft Excel, but they were frustrated by its computing and speed limitations.

On top of these challenges, Infopro Digital also knew they wanted to develop any new processes and skills within the company (as opposed to hiring third-party contractors) to keep costs and production delays low.

THE SOLUTION: DATA CENTRALIZATION AND DEMOCRATIZATION

Thanks to Dataiku, Infopro Digital's teams have been able to develop a solution that satisfies all of their requirements while also keeping ultimate control over the data in the hands of the IT team. With Dataiku, Infopro Digital developed a solution where:

- The IT team can make data available to analysts after only a few transformations thanks to reference datasheet sharing.
- From a technical standpoint, IT teams at Infopro Digital can take advantage of Dataiku's customizable capabilities and the ability to keep some freedom in terms of coding.
- CRM, accounting, and external datasets are consolidated thanks to Dataiku's focus on dataset centralization.
- Non-technical teams (like marketing) can build their skills and scale their efforts thanks to an intuitive, visual point-and-click interface. Longer term, the goal is to have them efficiently and independently leveraging website clickstreams and HDFS datasets.

Infopro Digital is now running one of the most advanced business-to-business advertising targeting engines on the market thanks to Dataiku.

- Data analyst teams can create reports and offer targeting samples within two hours, work that used to take them three days.
- Infopro Digital has been able to empower existing staff by internalizing the whole process.

“Dataiku really helped Infopro Digital keep a competitive advantage in the advertising market; we can now offer advanced targeting options in a snap. We now plan to expand our use of Dataiku to new markets and to our data business.”

Clément Delpirou, Chief Digital Officer at Infopro Digital



IV. MAIN CHALLENGES FOR AI IN MEDIA AND ENTERTAINMENT

ENSURING EFFECTIVE DATA ENGINEERING AND CREATING QUALITY DATA PIPELINES

Engineering challenges that we encounter within the M&E space are relatable for any industry that deals with large scale datasets — especially for data that includes any type of user generated content. Creating sustainable production pipelines to harness the relevant datasets is a challenge in data wrangling and is central to building the advanced analytics and tools that are used for decision making within the team.

DATA GOVERNANCE, PRIVACY AND REGULATION

Real challenges surrounding the treatment and ownership of personal information are spurring regulators around the world to catch up, and put pressure on companies to adapt to new privacy regimes. It is possible that issues relating to the safety and privacy of personal data will limit the ability of E&M companies to individualise the media experience going forward. Trust remains at a premium.

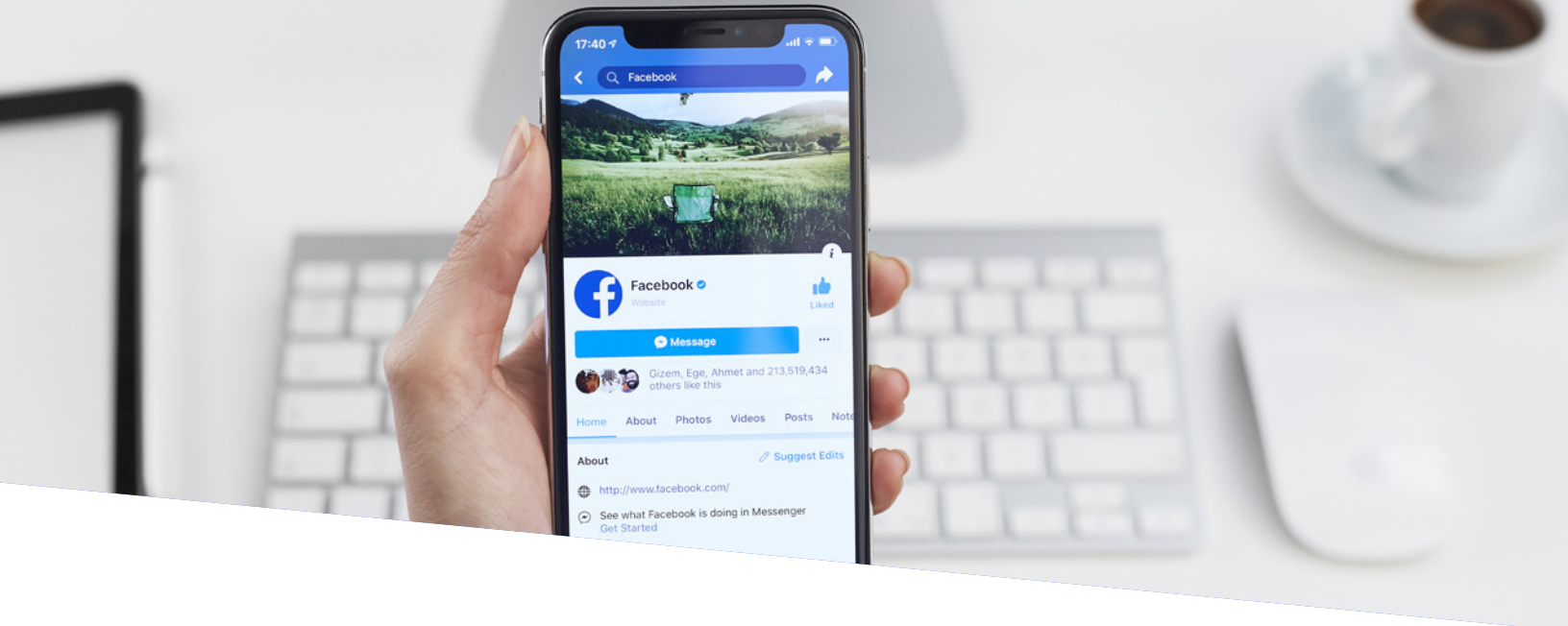
DATA PRIVACY MATTERS

When dealing with personal and/or potentially sensitive data, organizations should take it upon themselves to review any relevant regulations in their region(s) of operation and be prepared to comply with those regulations. The most important step in working with personal data is ensuring that the way consent is obtained for collecting that personal data is in line with regulations.

FURTHER READING:

Executing Privacy-Compliant Data Projects





SOCIAL MEDIA AND THE RISK OF MISINFORMATION

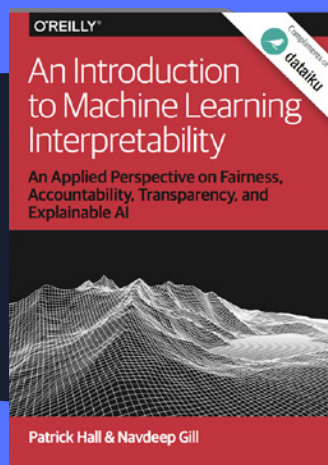
A Pew Research Center survey in 2018 showed that 20 percent of Americans reported getting their news often from social media, while only 16 percent reported relying on [print newspapers](#). Algorithms are a major part of every social media platform, and have an increasingly important role in news and media consumption. The information that individuals provide the platforms through their behavior feeds into algorithms that dictate what content will show up on their feed.

For media, and in particular news organizations, social media algorithms are of utmost importance, since they determine how many people are exposed to their content. Journalists have serious concerns about the impact that social media algorithms are having on news and the accompanying civic discourse. Because it is in the interest of social networks to drive “engagement” by showing users content that is likely to prompt a strong reaction from them, some worry they are favoring sensationalism or outright misinformation.

Therefore, it is of crucial importance for media organizations, whether they create or simply aggregate content, to have a deep understanding of how the algorithms favor and shape content, from the headlines to the topics.

FURTHER READING

O'Reilly Ebook: An Introduction to Machine Learning Interpretability



RECOMMENDATION ENGINES AND FILTER BUBBLES

Another challenge which could also increase the risk of media misinformation has to do with recommendation engines. While content recommendation is a core use case for media companies, it often creates so-called filter bubbles where a customer is only being given news stories and social media posts biased toward their existing beliefs, isolating users from differing viewpoints and perspectives. Filter bubbles might have a particularly damaging effect with regard to fake news. In order to avoid these problems, it is necessary to keep improving recommendation engines. By using a data science or AI platform, it becomes easier to map data sources, build and increase transparency around recommendations, as well as identify potential problems.

ATTRACTING THE RIGHT DATA TALENT

There are additional challenges that may be familiar to data scientists working in other industries. Just about every industry is trying to find data scientists and data engineers, but many companies hire by posting jobs that include a wish list of qualifications that is not necessary or required for the position and in many cases, the hiring manager doesn't even possess.

Media and entertainment started hiring for AI and ML positions a bit later than other industries and a lot of new data scientists compare it against companies like Google and Facebook without realizing they're comparing apples and oranges. Data science positions in media and entertainment are not equivalent to those offered in technology companies and while this may be a deterrent to some, there are a lot of unsolved problems in media and entertainment that curious and business result-driven data scientists can work on that are interesting and relatable.





V. UPCOMING TRENDS IN AI AND MACHINE LEARNING FOR M&E

IT'S ONLY GONNA GET MORE AND MORE PERSONAL

From traditional outlets, to digital media and online streaming services, to an ever-growing number of connected devices, media consumers are presented with a constant proliferation of channels and platforms. Experts have coined the term “ATAWAD” (no, this is not another Game of Thrones character you forgot about), which stands for “Any Time, Anywhere, Any Device” to describe the way content is consumed in this new playing field. More and more, consumers increasingly exert control not only over how and when they experience media, but also, over the content itself.

When it comes to responding to this ever-growing shift towards highly personal media experiences, in addition to continued advances in AI- and ML-based content personalization, there will be more integration across media experiences, such as digital video and traditional television, and widespread usage of digital assistants to aid content discovery.

DATA-DRIVEN CONVERGENCE BETWEEN TECHNOLOGY, TELECOM AND M&E PLAYERS

The biggest technology and telecommunications firms are now acquiring and integrating E&M assets at a faster pace and a larger scale than ever before. According to PwC’s TMT Deals Insights, the number of media and telecom deals rose 29% in 2017, to 876. Media and telecom M&As are expected to be driven by the continued convergence of media and technology, and in particular by the continued demand for data and artificial intelligence AI, [according to PwC](#).

MACHINE LEARNING AS A WAY TO TURBOCHARGE THE USER EXPERIENCE

ML-based recommendation systems are currently mainly used to suggest relevant content to consumers. But the recent advances being made in representation learning, which lies at the heart of neural networks, allow us to make better use of more unstructured data, such as image and text data, unlocking new, exciting use cases. As creators develop new content, embedding based recommender systems can autosuggest image assets for articles, for example, or help surface image assets without copyright limitations given a target image. Sequence-to-sequence learning can translate content from one language to another, but also from one writing style to another, allowing it to reach specific audiences more easily.





EVENTUALLY, AI WILL BE USED FOR THE CREATION OF CONTENT AND HELPING DRIVE IMMERSIVE VIRTUAL REALITY EXPERIENCES

We see AI itself becoming a new form of entertainment media through VR and AR technologies offering fully immersive experiences with intelligent avatars.

THE AI REVOLUTION POWERS A NEW GROWING DEMAND FOR TECH TALENT IN M&E

AI has opened up new opportunities within the media and

entertainment industry and created new roles in that require technical and quantitative skills. We can expect to see major job growth in this area over the coming years.

THE MEDIA AI DISRUPTION DOESN'T END WITH CONTENT

More and more, we'll see media companies leveraging automation and data science to improve the parts of business that are not industry-specific, such as supply-chain, human resources, accounting, security and employee recruitment and retention.

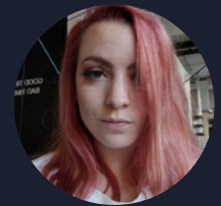


SPECIAL FEATURE: DATA STORYTELLING

*For this special bonus feature, we met with Alex Simonoff, Data Scientist at Spotify, the global leader in audio streaming, to tell us about the emerging trend of **data storytelling**.*

What exactly is data storytelling and how can M&E companies leverage it to understand their customers' needs and emotional drivers better to deliver better end-to-end user experiences?

Data storytelling is using data mining and domain knowledge to draw connections between trends in data to real-world outcomes or circumstances. Studying data to find friction points in the user experience (for example: how many clicks does it take on average for a user to complete a task or where in our app are users falling off the most) allows companies to innovate product to improve these experiences. By focusing on the customer's experience, data scientists are able to communicate with their teams the usage of their product or feature in a way that builds empathy towards customers. Using data storytelling to build empathy often results in more impactful work on the user experience being prioritized.



ALEX SIMONOFF

Data Scientist
Product Insights, NYC
Spotify





Do you think it's important for media and entertainment organizations today to study the cognitive relationships that people have with stories and storytelling and how they drive user behavior? How does data science fit into that?

In order for any organization, especially those focused on entertaining customers, to properly innovate their product they need to understand what customers actually want. One of the best ways to learn about customer needs and behaviors is by looking at their data and building a story around why they do what they do and how it relates to what they want. Data science is the best way to access that information, build the connections between data and experiences and recommend product changes to enhance these experiences.

What is the role of data storytelling in your day to day work at Spotify, as well as for Spotify's organization as a whole?

My job requires me to constantly tell stories with data. Everyone on a product team has their role and my role is to be a data specialist who not only understands what the data says but also translates those trends into takeaways and product recommendations. When an A/B test performs well (or poorly) I am digging into the data to find out why that might be. When we see a metric change over time, it's my job to be able to mine these changes to understand if there's a systemic issue or perhaps a particular audience driving this change.

At Spotify, data scientists all across the organization are working on building the best app features for playlisting, the best Premium offer, the smartest marketing plans and, like myself, the best ads and messaging mix for our Spotify Free users. Spotify embraces data storytelling in a way that allows data scientists to have an important seat at the table when it comes to deciding 'What's Next?'

CONCLUSION

Faced with increasingly uncertain markets, fierce digital competitors and the ever-growing amounts of data generated by media users everyday, media and entertainment organizations need ever-smarter tools to stay in the competition in the game.

When it comes to AI and data science implementation, media and entertainment has many of the same challenges as other industries: requiring effective data engineering, data quality and stewardship efforts, attention to data governance and data privacy, and usage of the most appropriate machine learning models for a given problem space. Some of the more specific to the media space challenges have to do with the increasingly important role that AI and ML algorithms play in the way people are exposed to media and get their information, as well as the potential misinformation and bias issues this could lead to. With some relevant expertise and education of people, careful governance processes, and the right tools, however, the technical, ethical and regulatory challenges are not impossible to surmount.

Data science, machine learning, and AI platforms are a clear win for media and entertainment organizations: they can provide a platform for organizations to optimize and deliver a next-level experience: an ability to understand their customers on granular level, deliver powerful, personalized content, user experience and storytelling, attract, retain and empower data talent, and create a new, sustainable business model. By choosing the right platform, M&E players can transform their business model and embark on the path to Enterprise AI to understand their customers and their businesses better, in order to deliver unique, differentiated, one-on-one experiences. Enterprise AI platforms open the door to drive media companies in this fast-paced world to new heights.





Your Path to Enterprise AI

Dataiku is the platform democratizing access to data and enabling enterprises to build their own path to AI. To make this vision of Enterprise AI a reality, Dataiku is the only platform on the market that provides one simple UI for data wrangling, mining, visualization, machine learning, and deployment based on a collaborative and team-based user interface accessible to anyone on a data team, from data scientist to beginner analyst.

300+
CUSTOMERS

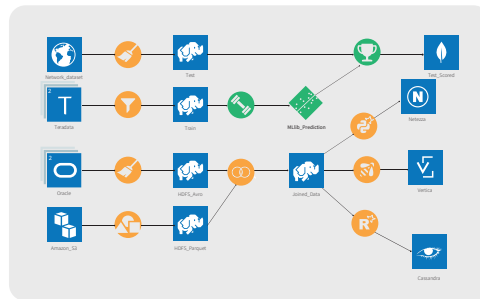
30,000+
ACTIVE USERS

*data scientists, analysts, engineers, & more

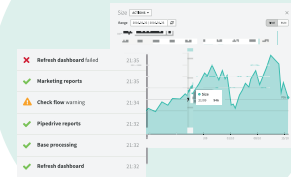


1. Clean & Wrangle

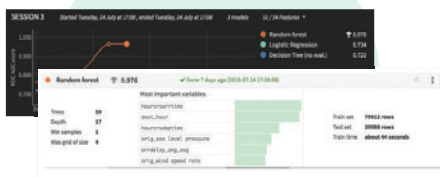
Name	Sex	Age
Robert W. Owen-Harris	male	22
Maria, M. James	male	28
Mohamed, F.	male	26
Alex, M.	male	30
MICaella	female	35
Hester, M.	female	29



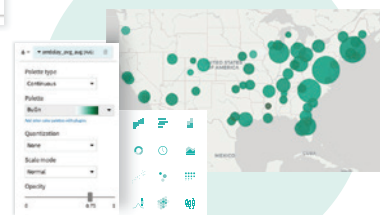
5. Monitor & Adjust



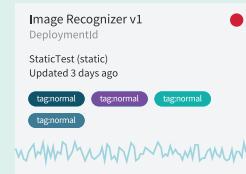
2. Build + Apply Machine Learning



3. Mining & Visualization



4. Deploy to production



WHITE PAPER

www.dataiku.com