O'REILLY®

Compliments of H2O.ai

The Future of Analytics

The New Landscape of Artificial Intelligence & Machine Learning Applications

Dan Darnell, Rafael Coss & Patrick Hall

REPORT

AI is Upside Down

With many AI projects failing to deliver value, we need to think differently. What if domain experts could partner with data scientists to build interactive AI applications that solve critical business problems and streamline customer experiences?

That is AI right side up – putting AI in the hands of business users across the organization and supporting customer interactions in every touchpoint.

H2O.ai AI-powered applications are a new way to use data to gain intelligence, get insights and predictions in an app that solves a business challenge.

Augment Data

H2O.ai augments and merges external data sets to help find answers faster.

Expertise

H2O.ai's data science experts are creating new AI apps that address business' most pressing problems.

A New AI Platform

The H2O.ai platform lets you augment, model, predict, and visualize your data.

Vertical Solutions

H2O.ai is working with customers and the community on building key apps with new models that solve specific challenges.

Provide real business value and get results faster with H2O.ai AI-powered applications.

Learn More >>

H2O.ai

9 781492 091738

The Future of Analytics

The New Landscape of Artificial Intelligence and Machine Learning Applications

Dan Darnell, Rafael Coss, and Patrick Hall

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

The Future of Analytics

by Dan Darnell, Rafael Coss, and Patrick Hall

Copyright © 2021 O'Reilly Media, Inc. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Rebecca Novack
Development Editor: Angela Rufino
Production Editor: Katherine Tozer
Copyeditor: nSight, Inc.
Proofreader: Abby Wheeler
Interior Designer: David Futato
Cover Designer: Randy Comer
Illustrator: Kate Dullea

October 2020: First Edition

Revision History for the First Edition

2020-10-16: First Release

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *The Future of Analytics*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

The views expressed in this work are those of the authors, and do not represent the publisher's views. While the publisher and the authors have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the authors disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

This work is part of a collaboration between O'Reilly and H2O.ai. See our statement of editorial independence.

978-1-492-09173-8

[LSI]

Table of Contents

Introduction . . . . **v**

1. The Converging World of Analytics **1**
Challenges in Current Analytics Projects 1
From Business Intelligence to Augmented Analytics 4
The Role of Automation in the Future of Analytics at Scale 5
The AI App Revolution 7
Current AI App Development Challenges 11

2. Modern AI Applications **15**
The Anatomy of a Modern AI Application 15
Detailed Application Examples for Key Industries and Functions 19
CPG Sales Forecasting with COVID-19 Data 21
Hospital Staffing Optimization (Healthcare Industry AI Apps) 22
Marketing Lead Optimization (Line of Business AI Applications) 23
Data Augmentation (AI Apps for Data Teams) 26

3. Case Studies: Real Impacts of AI Application in Business **29**
ArmadaHealth 29
Hortifrut 31
Jewelers Mutual 32

4. Adoption Challenges for Next-Generation Analytics. 35
Ineffective Data and AI Principles 35
Lax Security Practices 36
Inadequate Human Review 36
Downplaying Traditional Domain Expertise 37

Conclusion. 39

Introduction

In 2015, when I started writing *The Evolution of Analytics* with my colleagues Wen Phan and Katie Whitson, we made the case for machine learning in business. Five years later, it's time to make the case to use machine learning the right way in business. While I certainly don't know all the answers, a few themes stand out to me, looking back over the past five years. On the negative side, there's the endless hype about artificial intelligence (AI) and the tendency to deploy it in creepy and discriminatory ways. On the positive side, I see the growing government and public awareness of AI. I hope this awareness translates into the regulation of AI, improved interaction design in AI apps, and more corporate responsibility and governance for AI.

In this report, we'll introduce AI-driven applications that boost traditional data analytics with machine learning. While these apps may beat the odds, provide useful insights, and drive organizational value, such success stories don't serve as guarantees. In fact, everyone involved in organizational AI projects would be wise to take an inventory of AI's impacts on businesses, consumers, and the general public.

Another bright spot in AI over the last five years has been the development of technologies that increase human trust and understanding in machine learning.[1] These inventions have trans-

1 See the following examples: "This Looks Like That: Deep Learning for Interpretable Image Recognition", "Intelligible Models for HealthCare", "A Unified Approach to Interpreting Model Predictions", "Introducing AI Fairness 360", "Introducing Tensor-Flow Privacy", and the What-If Tool.

formed machine learning from a field of black-box algorithms to a field that is now capable of fierce debate around the concepts of algorithmic transparency, accountability, and fairness. This technological progress not only enables the nuts and bolts of regulatory oversight, but it also gives companies the power to govern their AI systems like the enterprise software assets they are. If you can block out the hype, you'll see that AI is really just software. And like all other enterprise IT resources, AI systems should be documented, managed, monitored, and governed.

Looking forward, I see successful AI deployments being aware of the risks of AI, taking on the associated governance burdens, and enabling humans to work together with computers to solve big problems. For businesses climbing to the next plateau in digital transformation, don't settle for any AI system. You'll need AI systems that are documented, transparent, managed, monitored, and minimally discriminatory. Moreover, these AI systems must support AI apps that are flexible, explainable, and, when appropriate, automatic. That's why Dan, Rafael, and I have written this new report, *The Future of Analytics*. It's a necessary update to the original *Evolution of Analytics* report, and we hope you find it to be a timely and useful guide through the new world of AI-powered analytics apps.

— *Patrick Hall*

CHAPTER 1
The Converging World of Analytics

In the broadest sense, analytics is the systematic analysis of data. This analysis makes the data consumable by people and systems, with the goal of understanding past outcomes and helping to predict future events. The adoption of analytics has driven a wave of digital transformation across industries where companies use data to power decision-making processes. Analytics projects, however, have not been without their drawbacks.

Challenges in Current Analytics Projects

Like many changes in business thinking, the first forays into data-driven decisions led down accessible but less useful paths. One such path was using dashboards to view historical trends to drive human insights from data, as shown in Figure 1-1. We now know that these traditional analytics dashboards alone can be insufficient to make better decisions as they provide only a historical summary.

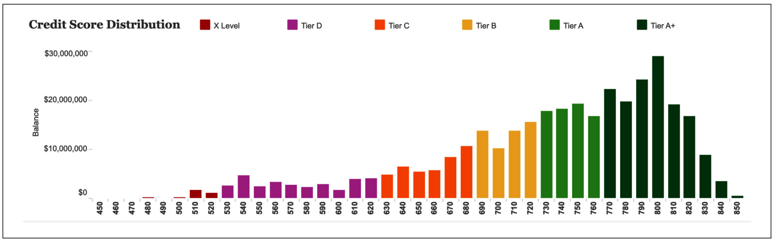

Figure 1-1. Typical dashboard with historical information

While historical trends are useful and can be predictive, they can also provide a false sense of confidence, when the future does not mirror the past. In the end, historical information and descriptive analytics alone leave business leaders to use their best judgment about trends in order to make decisions based on their own experience and limited view of the data. The result of this process is then highly dependent on the individual decision-maker's expertise, which yields highly variable outputs.

Using machines to find patterns in data and make predictions is another area of great promise for decision support. Machine learning models, trained on historical data, can look at new data and predict what is likely to happen. For example, credit card companies use machine learning models to determine who has access to credit and is likely to carry a balance on their credit card bill each month. Such models can even prescribe actions for users to take and recommend products or content of interest. This ability to predict future outcomes and prescribe actions has made machine learning a hot technology—and data science a hot profession.

For all its promise, machine learning has not reached widespread usage in production or within business applications where it can provide value and support business decisions. The challenges vary by organization and use case. However, the common themes in AI and machine learning adoption revolve around a few key areas, including a lack of resources, lack of business trust in models and their outputs, difficulty putting models into production and keeping them running, lack of consistent business involvement, and bottlenecks in putting predictive results into business applications. Let's discuss some of these dilemmas below:

Talent gap

"Data scientist" has become the hottest title and the hardest to fill position for many organizations. While many data analysts changed their titles to capture the new wave of interest, they lack the skills in languages like Python and R and an understanding of modeling algorithms and techniques needed for predictive model development. The lack of qualified data scientists in the market has put many companies on the slow track in AI development. Even with qualified data scientists on staff, companies find their output is limited due to the handcrafted nature of their work and the high maintenance costs of such models in production.

Model trust

Trust and the responsible use of machine learning models are evolving fields and the topics of entire books. What is clear is that for a business to use insights from predictive models, that organization must first trust the models' outputs. Trust can develop over time as users make decisions and begin to see that the model provides valuable insights. However, even before this can happen, data science teams must explain the models in business terms using techniques that show what factors contribute to each prediction, how the model makes such predictions, and how a model will behave on new data. Without a level of trust in a model, business users will not use predictive models to automate business decision processes.

Model operations

Another barrier to machine learning adoption is operational. Building an experiment to make predictions is one thing, but deploying a model into production is quite another. Again, this is a broad topic discussed in other books. At the summary level, models will need to undergo an extensive process of testing and validation, model packaging, and then production deployment and ongoing management for which many organizations are not prepared. Once these issues are exposed, models that once looked promising in the lab never find their way into production use.

Domain involvement

Probably the most overlooked barrier to AI adoption is the involvement and access by domain experts and end users. Too often, machine learning and AI projects take on lives of their own during development. While the business user may define a need or use case and even provide the budget for technology and data science teams, that is often the end of their involvement. Without a feedback loop between the domain experts and the development team, integrating AI into existing applications —or projects to create new applications—fails to get off the starting line.

Application development

All too often, even successfully deployed predictive models end up only as services. For predictive models to provide value to the business, these services must be incorporated into new or existing applications by application development teams.

> Without this application development step, the domain experts and end users can't provide feedback on the model because it is not in their context. This step also involves a new team of application developers in the project, and this team has a backlog of projects that, once again, can send your AI project to the dustbin.

To navigate the future of analytics, organizations must deal with each of these challenges. In this report, we will examine the solution to domain involvement and application development, as other reports and books cover trust, operations issues, and solutions.

From Business Intelligence to Augmented Analytics

Analytics means many things and covers a host of technologies that can sometimes seem quite different from one another. The two leading areas of analytics are business intelligence and artificial intelligence. Each of these fills a need for the systematic analysis of data, but they use entirely different means and typically support different user groups within the organization.

Business intelligence (BI) is the ability of an organization to use historical data analysis to understand business performance and ideally to improve business decisions and outcomes. Traditional business intelligence provides reporting technology for business users and analysts to view summarized historical information. Recent innovations in BI have added more real-time information to these reports, but the context remains to analyze available historical data. Common BI applications include key performance indicator (KPI) metrics reporting, executive dashboarding, and ad hoc data analysis.

Artificial intelligence (AI) is the ability for machines to perform tasks that people generally do, such as making inferences and decisions based on data, identifying trends, sorting images, responding to speech or text, and more. Traditionally, AI in the form of machine learning is used to create models that predict what will happen in the future based on past information. These models are used in batch processes to score or categorize records offline, or in real time to provide predictions in decision-making processes.

AI and BI are converging. Gartner, Inc., a leading IT analyst firm, views these parallel tracks converging and consolidating in support of a more unified analytics platform.[1] This convergence will drive better overall insights, understanding, and decision making, and overcome many of the challenges caused by using different analytics tools. In doing so, this consolidation will dramatically improve business agility and productivity, and create many opportunities for new and enhanced AI-driven business applications.

With the need to provide predictive and prescriptive insights, vendors across the analytics spectrum are changing their stripes. Companies that formerly focused on dashboards are incorporating predictive intelligence into their products. Similarly, companies that used to focus on data science and machine learning are offering analytic dashboards, including predictive insights. This convergence of business intelligence and data science with machine learning is called augmented analytics. Again, entire books exist on this topic. Many companies want to be the one-stop shop for descriptive and predictive analytics across the enterprise. However, simply combining different analytics styles may be exciting, but prove insufficient to meet business needs for actionable insights in a business context. Similarly, data science tools that help teams collaborate, manage, or automate some parts of the machine learning life cycle may also miss the mark. Indeed, the future of augmented analytics may prove to be something much different in the form of AI applications.

The Role of Automation in the Future of Analytics at Scale

If the future of analytics is one where AI reaches into every corner of business to enhance applications and processes, then automation plays a crucial role in achieving that vision. The traditional data science process involves weeks or even months of painstaking work to hand-code models using languages like Python or R and test various machine learning algorithms. This process can produce outstanding results, depending on the skill of the data scientist. Unfortunately, if AI is to scale, most organizations can afford neither the talent required for such artisanal work, or the time to wait for hundreds or thousands of such projects to complete.

1 "2020 Planning Guide for Business Analytics and Artificial Intelligence", Gartner Research, October 7, 2019.

The solution is the robust automation of the data science and machine learning process, known as automated machine learning (AutoML). AutoML does not replace the data scientist, but instead helps to automate the steps typically done by advanced data scientists. With AutoML, novice users can achieve robust modeling results, and expert data scientists can improve productivity by automating data science functions, including:

Feature engineering
: AutoML will identify and transform data into numeric forms that machine learning algorithms can understand. Advanced data scientists will also tease signal out of data by looking for the ways to combine input data columns together. AutoML should perform both functions, using the power of computing infrastructure to try different permutations of features to determine the ones that provide the best signal.

Algorithm selection
: For any given problem, there are many mathematical approaches that the data scientist could take. AutoML will typically look at the data and the type of problem, and then decide which algorithms are best suited. The AutoML will then run through these algorithms to see which ones provide the highest accuracy, most interpretable results, or the fastest scoring. The goals of the project will determine which model provides the best fit.

Parameter tuning
: For each algorithm, there are many different settings that will impact how the model is built and determine the accuracy and speed of predictions. AutoML can easily run through different tuning options to determine which parameter settings yield the best results, based on goals for the model.

AutoML should also include testing and validation of models to ensure high quality under various conditions and to avoid common pitfalls. The output of the AutoML process is a model that can be used for production scoring that includes all transformations to take in raw data and produce a score or response in a production environment. The automation of the data science and machine learning process allows more people to engage in the data science process from a variety of technical backgrounds, thus increasing the number of possible projects. Automation also dramatically decreases the time to complete projects and produce a production-ready model. These two factors lead to a dramatic increase in the number of AI projects that reach production.

Automation in production operations is also critical. Once deployed into a production environment, a predictive model will need to be monitored and maintained over time. When data patterns change, and the model is no longer able to predict as it once did, that model will need to be updated. Automation can play a crucial role in operations, allowing operations teams to monitor models without detailed knowledge of model techniques, algorithms, or modeling languages. Monitoring by exception, where guardrails are established around models to alert the operations team to issues, allows a relatively small group to manage a large number of models. This operations paradigm also removes the need to continually involve the data science team in production issues, which takes them away from building new models.

Automation of the end-to-end modeling and production process also lessens the dependence on specific individuals and makes the process more repeatable and robust. That is not to say that outstanding individuals cannot make outsized contributions, but rather that these contributions, like those of other users, must be captured in a repeatable framework. With this approach, when an individual is reassigned or leaves the organization, the process continues. Taking a systematic approach using automation across all data science and production projects leads to a kind of corporate memory, rather than individual memory, that allows the organization to scale up with thousands of AI-powered processes.

The result of such robust automation across the machine learning process is that models are not only easier and faster to develop, but once built, facilitates the automation of ongoing maintenance such as periodic retraining. This automation of updates is the key to building and maintaining hundreds or even thousands of AI-driven applications in an organization.

The AI App Revolution

Many in the analytics community believe that the last mile in AI is the production operations of models and endpoint services that deliver model results. You can see this in the way many vendors talk about consuming model results in batch processes or as an online service at the end of AI development. However, the truth is that no model can provide value—no matter how robustly implemented or how many computing nodes it's deployed to for scale and availability—if end users can't access it at the critical decision point, in context and in a way that they deem useful. This interface between the business user and AI happens in an AI-driven application.

Using AI in applications is not a new concept. Thinking about use cases and driving AI adoption from a business need should be the norm; however, as we have already seen, many such processes break down due to issues like trust, operations, and a lack of business context.

AI applications are the natural end product of a mature AI program, as outlined in Figure 1-2.

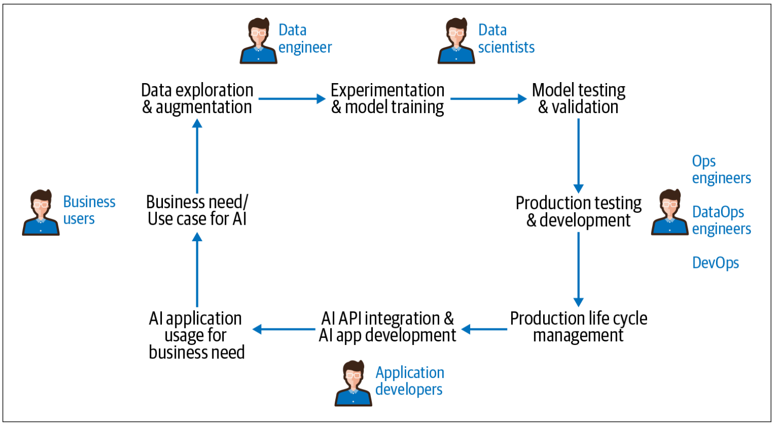

Figure 1-2. The AI application development life cycle

Each AI application starts as an idea from a domain expert or business unit with a problem to solve, or where AI is likely to improve efficiency or customer experiences. Although the terms are often used interchangeably, AI applications should not be confused with the business use cases related to them. The AI application is used by the business user to help aid or solve the problem from the use case. The process for developing AI applications can be broken down into the following steps:

Data
: The first question in developing an AI application always revolves around data. What data do we have and what data do we need for this use case? If internal data is insufficient, then

external data sources, either public or private, will be required to augment existing datasets. This set of tasks is often handled by data engineers who access data sources and create datasets using data management tools for use by the data science team.

Modeling
: With sufficient data, a data scientist can begin experimenting to develop a working model that can provide a useful prediction or prescriptive information for the use case. Without automation, this is an iterative process that can take weeks, or even months, to find the most accurate predictions that meet the business requirements and create a model that can be deployed in production.

Validation
: Once this model is developed and trained, it must be tested and validated, typically by a group outside of the data science team that developed it. These teams will use historical data to "back test" the model to prove that it would have generated acceptable results under a variety of conditions. Validation may also explore machine learning explanations to look for bias, to better understand how it makes the predictions, and to see any potential limits.

Operations
: If the model proves robust from a data science perspective, it will undergo further testing by the operations team. In these tests, the model will endure production loads on production systems to ensure that it can meet service level agreements for downstream applications. Once deployed, the operations team will manage and maintain the model.

Applications
: With a robust and reliable production model, application developers can now integrate those predictive insights into existing or new AI-driven applications. This process is best done with intense interaction with the domain experts to ensure that the application meets the business need.

This entire process is a cycle because as business needs change, new data becomes available and new algorithms or techniques emerge. Applications will need to be updated along with the machine-learning models.

AI-driven applications can take several forms, from enhancements to existing business applications to new, purpose-built decision support tools.

Enhancing existing app
: Currently, the most common AI-driven application augments an existing business application with AI insights. For example, showing a predicted trend line on an existing dashboard, adding a customer churn score (the probability the customer will churn), or recommending offers to an existing call center application. These projects provide value but can take application-specific development resources to integrate AI insights, making projects expensive, and taking many months to complete. Also, when adding AI insights into applications, those applications may need to be refactored or wholly redone to take full advantage of predictions or put the insights into context.

Purpose-built AI apps
: AI applications built from the ground up to support a specific decision or role are another way to deliver AI-driven insights. These applications include descriptive, predictive, and prescriptive insights focused on a particular problem. These purpose-built applications benefit from modern frameworks and focus on meeting immediate user needs. On the downside, application development resources are notoriously tricky to find, limiting the ability to develop these focused applications.

Prebuilt AI apps
: One solution to the time-consuming nature of application development is to purchase AI apps developed by specialists. These prebuilt apps can act as a springboard for the use of AI within an organization. The downside of such apps is that they have to be somewhat generic by design to accommodate any number of users. An example of a prebuilt AI app would be an application for automated clustering analysis. This application takes data as input and then outputs information and plots on clusters identified within the data. Such clusters of customers, for example, are useful in marketing to create segments and targeted campaigns. Prebuilt AI apps may also need to be customized to fit a particular organization's needs or data. When looking at prebuilt applications, business leaders should factor in the cost of such customization in their analysis.

AI applications are the value endpoint that many business leaders have been looking for to make sense of investments in AI and machine learning technologies and investments in data science teams. Incorporating AI or building new AI apps is how business users can finally see the value of smart models in their everyday work to improve customer experiences, increase employee productivity, and optimize business operations.

Current AI App Development Challenges

Incorporating AI into new and existing applications is perhaps more complicated than it seems. AI and machine learning predictions might seem like just another data source or another API service. To transform a business with AI, however, will take more than just putting trend lines on dashboards. Domain experts will need to partner with data scientists to build apps for their specific business needs quickly. These data scientists will also have to partner with application developers to create production apps. The apps they create will be interactive and help business users make smarter decisions. Understanding why existing tools and teams don't fulfill needs around AI apps is critical to understanding what organizations need to build the right team and adopt the right technology.

The Impedance Mismatch Between AI and Existing Web Frameworks and Teams

Software web development frameworks for building business applications are well established. When you want to build a web application, developers will use web frameworks and languages like Django, Flask, Express, Ruby on Rails, HTML, JavaScript, Spring, and more. For business applications, developers may default to classic BI reporting and dashboard tools or others. To engage these teams in new AI application development creates new issues. First, application developers and web developers are already swamped with projects. Bringing them a new project that involves AI will likely put the project at the end of the queue in favor of projects that can be delivered with existing tools. Then there is the impedance mismatch between the DevOps and MLOps life cycles. The lack of knowledge about AI and machine learning also means that developers might not understand what the application is doing; why, when, and how the model should be applied; how to interpret the results; and when to retrain a model, which can lead to delays and expensive rework.

Web development is not AI-friendly

While applications using AI may seem like any other web application, there are key differences that will doom any project conceived and executed using AI and web frameworks. First, many web frameworks are designed for simple data interaction patterns. Typical web data applications tend to be transaction-focused with clicks and simple inputs driving subsequent changes in pages to provide new experiences. These applications tend to work with web frameworks that know how to interact with transactional or operational databases. The analytics in these applications often only work with the operational database and not the analytics databases (enterprise data warehouses), and therefore have minimal analytics.

Generic AI Apps

One approach to AI apps is to build generic apps that allow users to input values and see results from predictive models. These apps have names like What-if, Optimizer, and Predictor. These applications are little more than a front end to the model API. While such applications can help business users interact with predictions, these applications are not tailored to a business need or user. The result is that such generic applications become little more than a way for business users to test models and become comfortable with their outputs. In addition, generic applications are unlikely to be designed for specific needs like peak loads in a call center or for online shopping. These applications also have static workflows, so there is minimal additional exploration, simulation, and evaluation.

Noninteractive Experience

Many analytics applications are used to pull data from a data warehouse to generate reports or dashboards. Interactivity with such data is limited to filtering and slicing the available data. In this framework, additional requests will pull a new set of data. As the user looks at dashboards, they expect to see the latest data and to have the data visualization dynamically update as new data arrives, but this is not the case. Even for IoT applications that process data in real-time streams, the user's dashboard will not update unless the user specifically requests it. For AI app development, systems should have both push and pull capabilities to provide an interactive experience for the user, which should also include real-time predictions in context.

Scarce Development Resources

In many organizations, application developers are as scarce and valuable a commodity as data scientists. As mentioned earlier, getting a project into the queue of IT app developers can postpone a project by months. When thinking about AI application development, organizations should consider a two-stage approach that includes rapid prototyping and final production development. This rapid prototyping phase should be managed by analytics or data science resources who can work closely with the domain experts in the business. Once an AI application design is completed and has shown business potential, only then does the application development team become involved.

The world of analytics is changing. With business intelligence and artificial intelligence merging into augmented analytics, the landscape is ready for a new set of technologies and applications for building AI into applications. Automation of AI model development, training, and operations play a critical role, but focusing on the traditional data science life cycle or leveraging automation in AutoML alone is insufficient to provide better customer experiences and smarter decisions for business users. The next step for analytics is the creation of AI applications. These new AI applications will drive new levels of business value by bringing descriptive, predictive, and prescriptive intelligence to users across the enterprise.

CHAPTER 2

Modern AI Applications

For AI to have a serious impact on organizations, users of all types—from the manufacturing floor to the executive suite—must have access to AI to make better decisions, improve customer experiences, and optimize business operations. This ubiquitous access to AI will come in the form of applications used to make everyday decisions and drive processes. These AI-driven applications will change the way employees do their jobs and how companies interact with their customers and partners. In this chapter, we'll explore the makeup of such applications and look at detailed examples of applications focused on specific business problems.

The Anatomy of a Modern AI Application

The modern AI application is the product of augmented analytics, automation, and application development. To better understand how the AI application differs from traditional applications and dashboards, let's dive into the various components and key features of a modern AI application (see Figure 2-1):

Specific
: Users of AI applications typically have specific goals and decisions in mind. The interface of the application is designed to support those specific decisions, such as making hospital staffing assignments, as shown in Figure 2-1. Unlike a dashboard, however, the interface is interactive; the user inputs questions or parameters, and then interacts with the analysis to further refine the results or to ask additional questions.

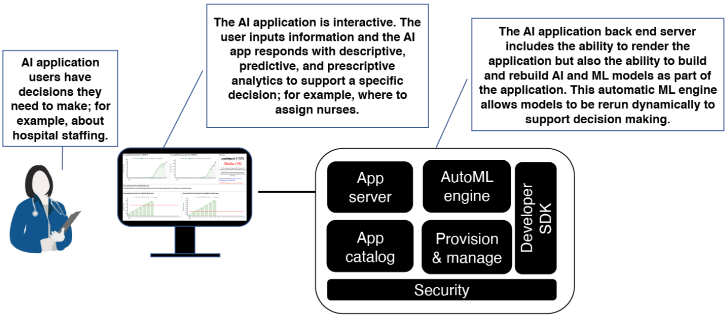

Figure 2-1. Example of a modern interactive AI application

Interactive

Modern AI applications are also push-driven versus pull-driven. Many legacy BI dashboards pull data on request or periodically from a database. The information displayed is correct as of the last refresh. Modern AI apps are push-driven because the interface is continuously being updated as data arrives and is updated along with the latest predictions, creating a more dynamic experience for the user and ensuring that they have the latest information.

Automated

Modern AI applications include automation to maintain their performance over time. A critical function in modern AI applications is the ability to build and rebuild AI and ML models to support new data or changes in request parameters. An open and extensible AutoML engine is used by the developer or data scientist involved in the AI app project to build models used in the app. Using AutoML, versus a hand-coded approach, is desirable for the business and the users as it allows the application to be updated without needing resources from the data science team, which are typically busy with other projects. The use of AutoML also creates a well-documented and repeatable process so that applications don't break down as they are updated.

Key Components of an AI Application Platform

Modern AI applications have a variety of components that facilitate interactivity and allow for easy updates and management:

End user application interface
: The business user will access AI applications either within existing business applications or as standalone applications. End users will access applications through a browser with a specific URL or through an application catalog that showcases available apps. The interface consists of a series of steps to collect data and processing directions. This workflow guides the user through the steps required to create the outputs they are looking for.

Development environment
: Application developers and data scientists will have their own interface to develop apps with a software development kit (SDK). Ideally, this development environment can be deployed to their development environment or laptop where they can iterate through application changes. Developers can use this environment to create rapid prototypes and to share these prototypes with domain experts for review. Once the application is ready to be shared with the user community, the developer can publish the app to the server environment for sharing.

Application server
: The application server provides the needed services and scalability for production applications. Once applications are published and access controls are set, then end users can see applications that are available to them.

Application catalog
: If the user has access to more than one application, then an application catalog capability is required so that the user can access, load, and run multiple applications, as shown in Figure 2-2. Using this functionality, the administrator can also load new applications for the users. When each user logs into the system, they will see the applications available to them.

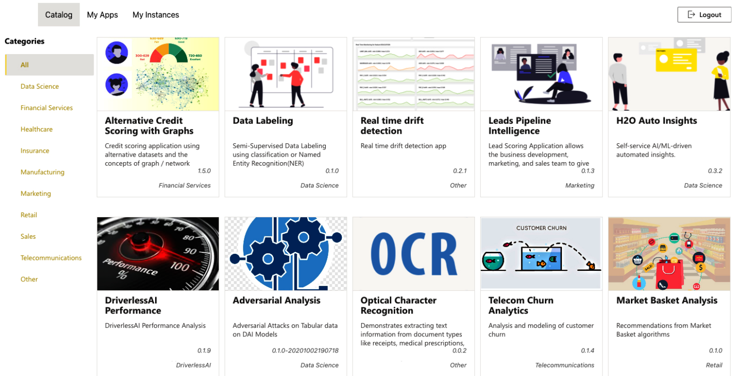

Figure 2-2. Application catalog gives users in an organization access to available apps

Security

Among the many functions of the application environment, the security of the data and application is highly important. In this system, AI applications should only be accessed by designated users and updated by administrators. As data travels between the application server and the user's computer, the information should be encrypted to ensure that it cannot be manipulated in transit. This is especially true when personal or medical information is present.

Provisioning and management

AI applications are deployed from the server for the use of each user. The model is similar to "app stores" that mobile phone users are used to. This is different from shared software, where multiple users share the same application. This separation allows the users to customize the apps to meet their needs and keep those settings for future use. The application management system then has to provision the application for the user to consume and track who is using the app and what version they have downloaded. Application updates are then made available for those users to update their app to gain access to new functionality. Administrators may also choose to update applications directly in the event of a security concern or other serious issue.

The framework and key components of modern AI applications allow for usage in a variety of use cases to create specific, interactive applications across industries.

Detailed Application Examples for Key Industries and Functions

AI applications may seem abstract to business leaders. The following examples show how AI applications are used across industries and lines of business functions, especially with business changes due to COVID-19.

Mortgage Lending (Financial Services AI Applications)

In early 2020, COVID-19 shutdowns caused massive unemployment and other societal disruptions. The models that banks had previously used to determine mortgage eligibility were suddenly ineffective due to a dramatic change in conditions.

To better understand potential borrowers in these new conditions, a leading US bank partnered with an AI technology company to develop a new mortgage lending application that would take into account COVID-19 related data, including infections and unemployment. The new application centers around scenario planning and gives the user the ability to input various factors and review possible scenarios for risk (Figure 2-3).

The output is a series of scenarios that are specifically designed for the mortgage lender to understand what happens with changes in data. By reviewing the scenarios, the bank now has a better understanding of the factors that will lead to defaults under changing conditions.

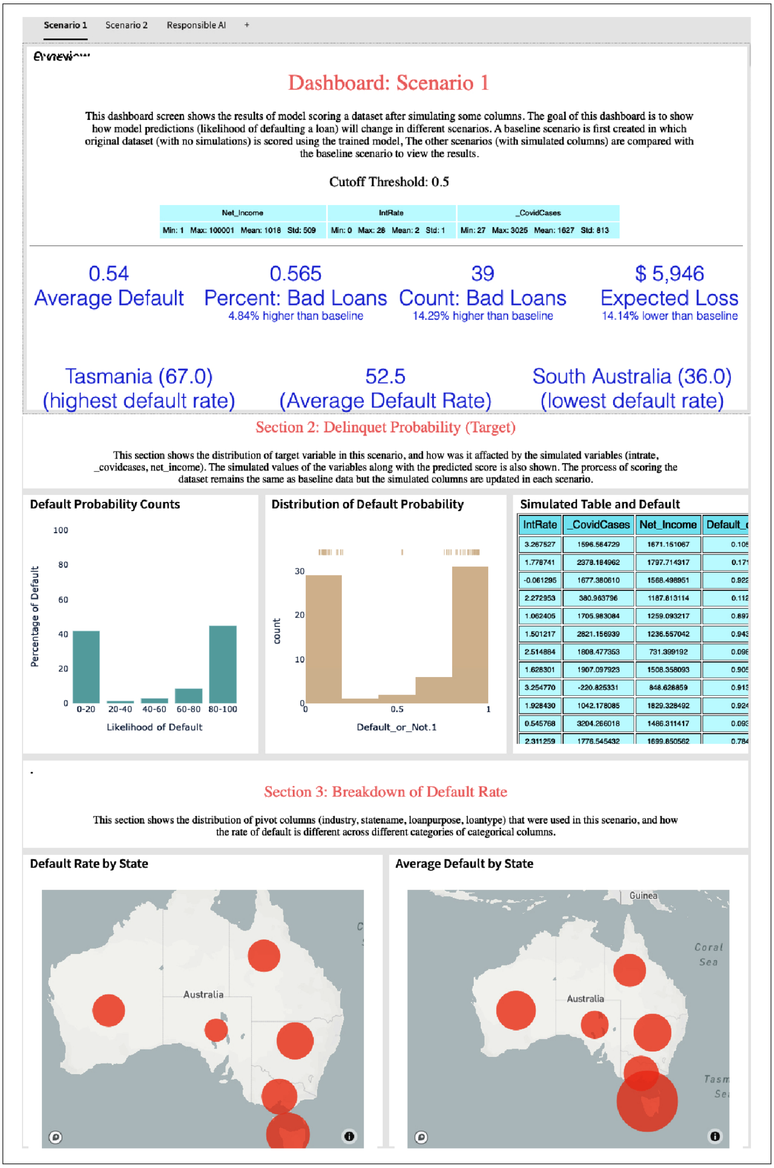

Figure 2-3. Mortgage lending AI application using localized unemployment and other data

CPG Sales Forecasting with COVID-19 Data

When the shutdowns started, many businesses found that years of experience and data were unable to predict future demand. For consumer packaged goods companies, this was especially true as demand for commodities like cleaning supplies and toilet paper changed overnight. Forecasting models built on typical seasonal demand and regional patterns could not predict demand under a new set of conditions based on COVID-19.

A leading European CPG company with a large number of products was quickly put under pressure as their products went out of stock across geographies worldwide. To replenish supplies, the manufacturer needed to know which products would be most in demand and where. Working with a leading AI technology company, they developed an AI application to predict demand for each SKU at each retailer and region in a matter of days (Figure 2-4).

This new application used an updated forecasting model that enriched historical patterns with new data sources to predict the new demand patterns. A new predictive model alone would have been insufficient. What was also needed was an interactive application that business leaders could use for scenario planning.

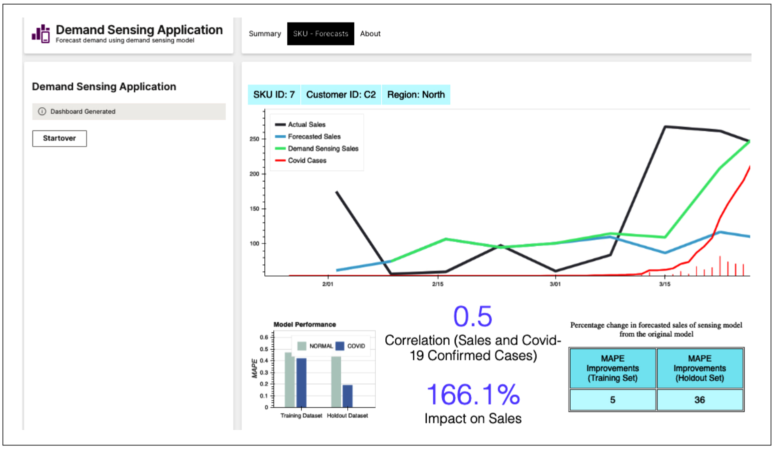

Figure 2-4. CPG sales forecasting AI application using COVID-19 data

The demand-forecasting AI app allowed the business leader to input the products, regions, and retailer partners of interest to see the predicted demand over the forecast period to plan for production and delivery of products across geographies.

Hospital Staffing Optimization (Healthcare Industry AI Apps)

Faced with COVID-19, US hospitals suddenly experienced potential shortages in resources, including ventilators, personal protective equipment, and nursing staff to fight the epidemic. Under normal circumstances, forecasting models were not really required because needs for equipment and staff were relative to the size of the hospital and service area. With a new and deadly disease, however, resources would be needed based on a new set of factors including infection rates and demographics.

A leading US integrated healthcare company, serving over twelve million people, realized in early March that they did not know where to place resources across their hundreds of hospitals and medical offices. They partnered with a leading AI technology company to build a new AI application to forecast needs for ventilators and nursing staff across locations.

The new AI application allowed users to input geographic areas, hospitals or clinics of interest and then see how infection rates and demographics for their patient community would impact resource needs (Figure 2-5).

Using the new app, the administrative team was able to allocate resources to the sites with the most need and get people and equipment there ahead of outbreaks in cases, which helped save lives for both patients and healthcare workers.

Another critical area for AI apps is at the line of business level, including key business functions like marketing, sales, services, and support.

Figure 2-5. Hospital resource forecasting AI application using COVID-19 data

Marketing Lead Optimization (Line of Business AI Applications)

Marketing departments in B2B companies generate leads from marketing programs that are followed-up by the sales department. Not all leads, however, are of equal importance. For a given company, or even at a given time of year, the factors that make a good lead could change. Traditional lead scoring is done by rules based on the items of content that a prospective customer views or downloads. A customer who comes to a live webinar might get more points than someone who just downloads a solution brief, and so on. Once a lead is passed to sales, each salesperson uses their own experience and bias to determine which leads to follow up with. The problem with this method is that it relies on human intuition to determine what makes a good lead and is often "one size," leading to missed opportunities.

The lead scoring application allows the business development, marketing, and sales team to predict which prospects or accounts are likely to make a purchase in a given period. A machine learning model looks at the history of previous prospects and the interactions with the company to build a predictive model. This model is then used to score new prospects. The higher the score, the more likely a prospect will be to engage with sales and spend money on products in a given time period.

A more granular ML lead scoring approach can impact many areas of the business, from marketing to sales. Knowing which leads have the highest potential helps everyone prioritize where to spend their time. Marketing knows exactly what leads to pass along to sales and which leads they need to nurture further. Sales knows which meetings to prioritize based on the relative score.

The lead scoring application uses AutoML to produce a model to predict which leads in the funnel are good or bad. AutoML trains and tunes several models based on the training data provided. To run the lead optimization app, the user first selects the settings for the geographic regions and data sources they would like to use (Figure 2-6).

The output of the application depends on the settings the user selected. Figure 2-7 shows a collection of reports that the user has to explore the data about the leads, both good and bad. This analysis allows the user to understand what factors are important to create good leads and what contributes to poor leads. Marketing users can then use this to tune marketing campaigns to those areas that produce better results. Sales teams can use this information to find other potential customers with similar profiles. Depending on the needs of the users, the application can also produce a list of scored leads that can be used directly for lead follow-up.

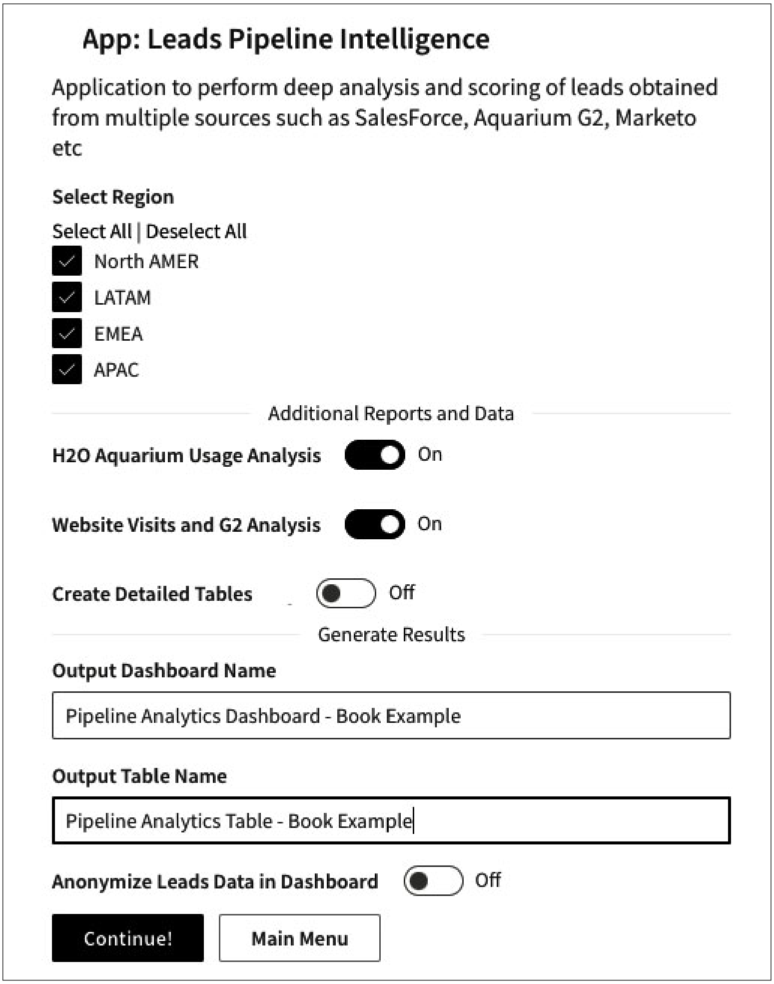

Figure 2-6. Users adjust the settings to produce their desired analysis

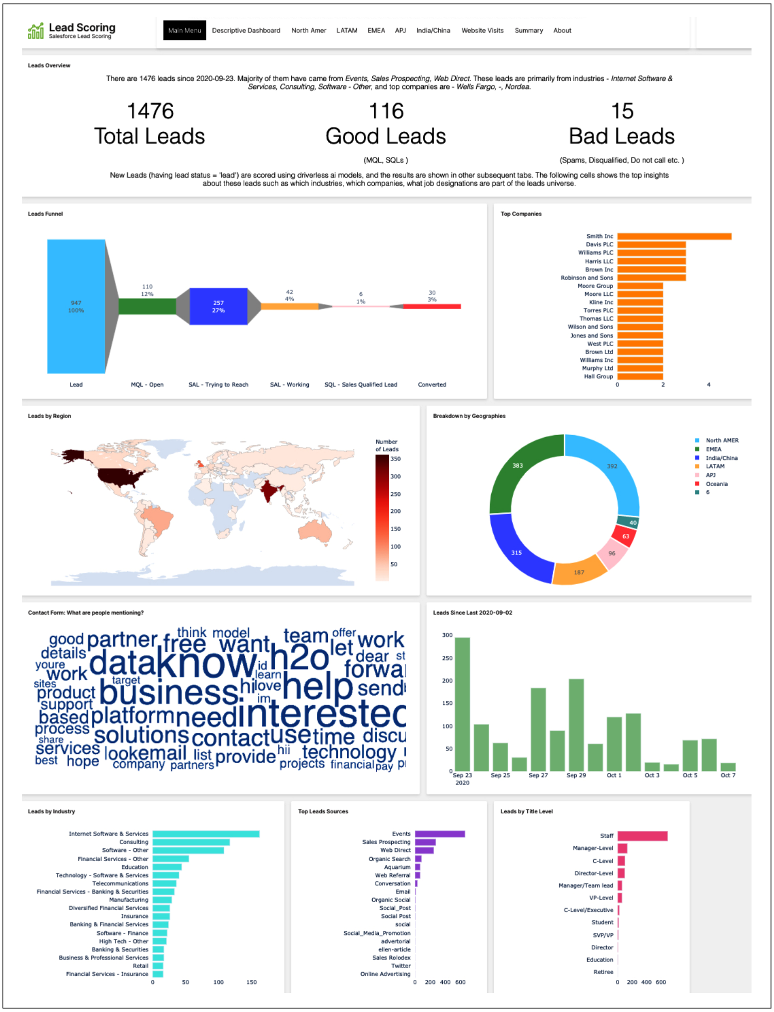

Figure 2-7. Lead dashboard provides detailed information about what makes a good lead

Data Augmentation (AI Apps for Data Teams)

Another key step in the development of any analytic process is access to data. When an organization's data is insufficient or when additional information could improve model accuracy or insight, then external data sources will need to be acquired and integrated with existing data. With the onset of COVID-19, many companies found that existing predictive models no longer provided accurate results. Rather than look for more of the same data, i.e., a longer dataset, some organizations looked to create wide datasets by adding new data from external sources. With traditional processes, finding datasets and discovering how to merge them, and then teasing these signals out of the data can take months of painstaking development and testing.

The data augmentation AI application uses fuzzy matching techniques to find relationships across datasets. This matching algorithm saves time by finding new related data quickly using information like time, location, or names in the data. Included in the application are public datasets, like US census data, crime data, housing data, and so on. The application also includes samples of private datasets for users to try. With access to these datasets and an easy way to match them to private data, organizations can quickly determine if public or private data can provide additional signals to improve model accuracy or provide insights to business users.

Let's explore the data augmentation workflow as part of the lending app (Figure 2-8). The process begins with selecting the initial dataset for data augmentation. Users then select what datasets they would like to consider for augmentations and the method for augmentation (automatic or guided). Next, users are presented with a variety of datasets that have matching columns that can be merged with the first dataset.

The original datasets are then automatically blended, and more features are added. This newly enriched data can then be used to do further analysis and potentially build better machine learning models. This automated enrichment process allows data analysts and data scientists to quickly and easily explore public and private datasets to see if they provide new insights for their use case.

When data augmentation is combined with AutoML, data scientists can very quickly see if new data provide valuable signals by creating new models and comparing them to existing models. Without this ability to rapidly and cheaply build models, exploring new datasets for small increases in signal is not cost-effective.

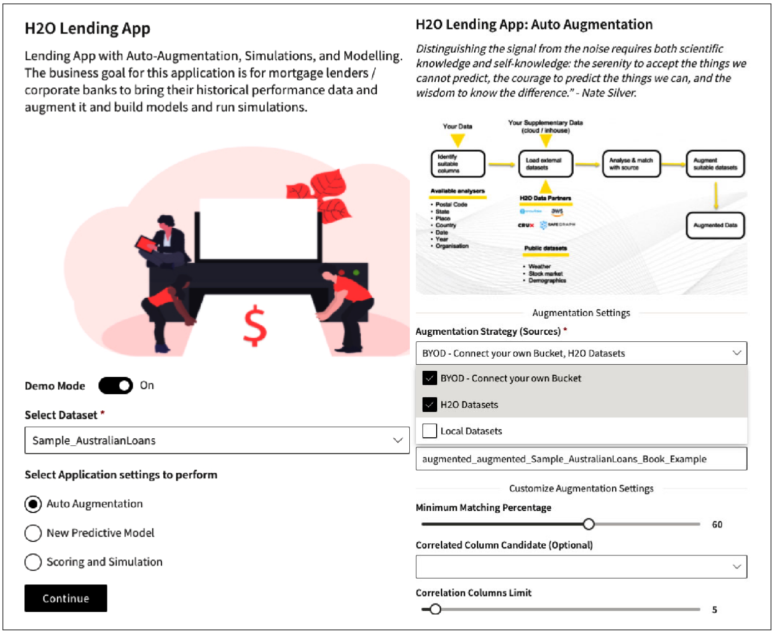

Figure 2-8. Workflow guides the user to select data augmentation level and augmentation datasets

CHAPTER 3

Case Studies: Real Impacts of AI Application in Business

AI and machine learning can have significant impacts on business transformation. Many companies and platforms use AI to create solutions that have a countless degree of impact on businesses. In this chapter, we are going to focus on just a few of these solutions, and how H2O.ai has worked with companies to solve their problems.

ArmadaHealth

ArmadaHealth is a health data science and services company founded to help people access the right physician or expert for them. Their unique solution, QualityCare ConnectSM, combines big data and expert clinical insights which points straight to the root cause of healthcare access problems. ArmadaHealth does this by applying sentiment analysis on customer reviews and advanced analysis of experts' wisdom to understand the consumers, objectively finding providers that meet their needs and preferences, preparing them, and delivering timely access to a choice of the most appropriate physicians for their condition (see Figure 3-1).

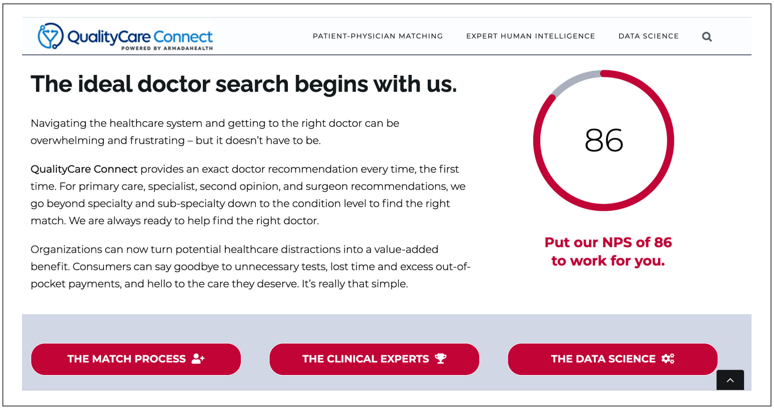

Figure 3-1. The QualityCare Connect app on the ArmadaHealth website

Challenges

Finding the right specialist is the first step to receiving the right care. However, consumers are often not equipped to navigate the complex and confusing healthcare system. It can be challenging for patients to discover which specialist they should approach for different health situations and, even with a referral from a primary physician, it can still be a long process to find the right specialist who can accurately treat them while also providing a satisfactory patient experience. Finding the right match between patient and doctor quickly can solve major problems and save lives.

Solution

AutoML is an essential part of reaching ArmadaHealth's goal of delivering accurate patient-expert matches through their online application. Using the H2O.ai platform, including automatic machine learning, the company is able to build and train a natural language processing (NLP) model to identify the sentiment (positive, negative, neutral) in each customer review. The company looks at three main aspects in each review: treatment outcome, communication, and attitude. These three aspects are critical to finding the expert that best matches customer preferences.

Results

ArmadaHealth has achieved its goal of helping people access the right physician or expert for them using a purpose-built AI app. Among the various positive results of using AI, ArmadaHealth saw increased net promoter scores and faster model building and deployment.

Hortifrut

Hortifrut, based in Chile, is the largest producer of blueberries in the world, and operates farms in Peru, Chile, Mexico, Argentina, the United States, Spain, Morocco, and China, with distribution across 37 countries. Hortifrut holds 25% of the world's blueberry market and uses AI to make distribution decisions across their expansive operations. They are able to predict the quality of the blueberries from origin to final destination, improving the consumer experience with higher quality products, and increasing revenue throughout the supply chain.

Challenges

Transporting fruit from the farm can take weeks, so Hortifrut has to predict the quality of produce upon arrival. Not being able to do this accurately can impact customer experience and revenue loss. But getting such predictions accurately can be a difficult task, given the complexity of the distribution channel, weather data, variety of datasets, shipping times, and more. Traditional machine learning methods and toolkits took months to build accurate predictions and production-ready models. To scale the use of AI under these conditions would require hiring additional data science talent and increasing the budget.

Solution

Hortifrut leveraged the H2O.ai platform to have better predictive insights into the quality of their blueberries. They used capabilities such as feature engineering, NLP, explainability, time-series analysis, visualization, and scoring pipelines. Hortifrut is now able to scale their data science efforts in order to deliver use cases such as predicting the quality of blueberries based on features like variety, farm origin, shipping time, vessel, and packaging, without hiring additional data science talent in the team. These results are delivered directly to business users making decisions, so they can take the correct actions for each shipment.

Results

Hortifrut achieved the following key benefits using AI apps:

- Hortifrut has saved a significant amount of money by reducing perishable claims. If the berries are spoiled at their destination, there can be a loss of revenue from customer claims and also cost valuable customer satisfaction.
- Hortifrut has been able to deliver real business results with a small data science team by improving the productivity of the team instead of hiring more people.
- Hortifrut is able to reduce the model development time from three to five months down to three to five weeks.

Jewelers Mutual

Jewelers Mutual is one of the United States' and Canada's most established and trusted providers of affordable and comprehensive insurance for jewelers and consumers. As a leader in driving customer-focused innovation and providing the latest technology to a long-standing industry, Jewelers Mutual uses AI to deliver exceptional customer experiences, prevent losses, and provide better protection and policies for both jewelers and consumers.

Challenges

The leadership at Jewelers Mutual recognized the need to invest in analytics, AI, and machine learning for improving overall customer experiences. Their business relies on both being able to effectively protect their customers' businesses, and providing personal insurance directly to consumers—both with innovative customer experiences. Jewelers Mutual has been at the bleeding edge in adopting AI. They collected data already available from losses, customers, and multiple other sources, which weren't tapped into before.

Solution

Jewelers Mutual standardized on the H2O.ai platform to develop predictive models. Their first deployed AI application helped commercial underwriters understand their customers better and provided the reason codes as to why decisions were made by machine learning models. These insights were then made available through an app to the underwriters. Having interpretability and explainability as part of AI model development and deployment was also instrumental in convincing business stakeholders to use AI to make business decisions.

Results

The success that Jewelers Mutual has seen in adopting AI in their business is a testament to the fact that regulated industries can achieve real competitive advantage using AI. For example:

- Jewelers Mutual has been able to offer more competitive jewelry protection insurance rates to its customers.
- AI insights are leading to some interesting outcomes; during the 2019 California wildfires and power outages, the Jewelers Mutual team was able to identify jewelers that would need additional physical security personnel to protect their properties and inventory.

CHAPTER 4

Adoption Challenges for Next-Generation Analytics

The new world of AI apps promises to make AI available to everyone across organizations for every function, from frontline employees to executives. Before jumping into this AI-powered revolution, there are a few critical issues that organizations should consider as they invest. In addition to widely-discussed staffing and technology issues, AI is presenting several less well-known challenges that will be the focus of this chapter.

Ineffective Data and AI Principles

According to AlgorithmWatch, dozens of organizations, including governments, have published data and AI usage principles.[1] These principles attempt to set broad guidelines for the use of data and AI within an organization and also signal to the organization's peers, competitors, employees, or customers that they are considering certain data and AI risks. Unfortunately, AlgorithmWatch also recently published a report stating that only 10 of 160 reviewed sets of principles were enforceable.[2] So, if your organization does create data and AI principles, keep in mind that a major pitfall to avoid is

1 See AlgorithmWatch's "AI Ethics Guidelines Global Inventory".

2 See AlgorithmWatch's "In the Realm of Paper Tigers: Exploring the Failings of AI Ethics Guidelines" by Leonard Haas and Sebastian Gießler, with additional research by Veronika Thiel, April 28, 2020.

ineffectiveness. Getting a technological perspective, along with ethical, legal, oversight, and leadership perspectives into organizational AI and data principles are perhaps the best way to avoid such issues.

Lax Security Practices

In the hype around machine learning and data science, and in a rush to build end-to-end data and AI products (the kind of technology that reaches into data centers, out to the public, and back into data centers), data scientists can be intentionally or accidentally given too many privileges in an IT system. This is an ethics and security problem. If the same person can manipulate a database, create a predictive model, and make it operational, they can make a predictive model do what they want it to do, and in very subtle ways. (Maybe it's to give their girlfriend's mother a giant loan, or to deny loans to people in a political or socially discriminatory way.) Regardless, these kinds of insider attacks against AI can cost your organization money and are another avenue by which discrimination can enter into AI. More standard concerns about data privacy must also be recognized by AI practitioners to ensure solid security. Hence, AI systems and the teams working on them should be under the same, if not stronger, security constraints as other employees.

Inadequate Human Review

Related to the practice of model risk management, the concept of effective challenge is used to improve AI implementation at large financial services organizations in the US. An interpretation of an effective challenge is that, when building AI systems, one of the best ways to guarantee good results is to actively challenge and review each step of the development process. Of course, a culture of effective challenge must apply to everyone developing an AI system, even so-called "rock-star" engineers and data scientists. For instance, the Federal Reserve System's famous SR 11-7 guidance on model risk management makes no exceptions for rock-star data scientists, and there's probably a good reason for that: a rigorous human review of AI systems is one of the best known methods for mitigating risks associated with AI. One easy way to start to build a culture of effective challenge is to hold mandatory weekly meetings where alternative design and implementation choices for AI systems are put forward, questioned, and discussed.

Downplaying Traditional Domain Expertise

Real-world success in AI often requires input from humans with a deep understanding of the given problem domain. Of course, such experts can help with feature selection and engineering, and interpretation of AI system outputs. But they can also serve as a basic sanity and usefulness check mechanism. For instance, if you're developing a medical ML system, you should consult with physicians and other medical professionals. How will generalist data scientists be able to understand the subtlety and complexity inherent in medical data and the results of AI systems trained on such data? They might not be able to, and this can lead to AI incidents when the system is deployed. The social sciences deserve a special callout in this regard as well. Sometimes called "tech's quiet colonization of the social sciences," technology companies are pursuing AI projects that either replace decisions that trained social scientists should make, or use practices, such as facial recognition, for criminal risk assessments that have been highly criticized by social scientists.[3]

AI Security and Privacy

Like nearly every other powerful commercial technology, AI systems are subject to failures and attacks. These can include the kind of hacks that plague other public-facing IT systems, wherein attackers block services with massive amounts of incoming web traffic or insert themselves between an AI service and a consumer. These can also include specialized concerns regarding training and output data privacy and security, or even highly specialized attacks on underlying machine learning algorithms.

In terms of data privacy and security, there are traditional data security concerns related to the confidentiality, integrity, and availability of input training data, intermediate data generated by the AI system, and output response data from the AI system, but there are also increasing legal and regulatory obligations around data privacy. These can include everything from the legal basis for data collection, to anonymization requirements, data retention limitations, and

3 See, for example, "To Really 'Disrupt,' Tech Needs to Listen to Actual Researchers", *Wired*, June 26, 2019; Rumman Chowdhury's post on Twitter; "AI Researchers Say Scientific Publishers Help Perpetuate Racist Algorithms", *MIT Technology Review*, June 23, 2020.

alignment with organizational privacy policies. Moreover, security and privacy breaches can also trigger breach reporting requirements. Because AI is so hungry for data, all of these can indirectly, or even directly, impact an organization's use of AI.

For specialized attacks against machine learning algorithms that underpin most of today's AI systems, organizations should have several known attack vectors on their radar, including:

- Insider manipulation of training data (i.e., "data poisoning")
- Manipulation of model outcomes by external adversaries
- The theft of intellectual property, like models and data, by external adversaries
- Trojan horse code or manipulations buried in complex machine learning software and related artifacts, like model weights that give a favorable outcome under certain conditions only known to hackers or external adversaries

While basic security practices are an effective shield against some attacks on machine learning, it's important to consider these attacks as part of updated model risk management or information security policies. Organizations can also leverage security audits, bug bounties, and red-teaming to help understand their vulnerabilities and to fortify their defensive measures.

The future of analytics in the enterprise is bright, but before you begin or scale up on your journey to build AI applications across the enterprise, there are clearly some organizational and security issues to consider and mitigate before taking the plunge.

Conclusion

The future of analytics is sure to take many forms. With the convergence of traditional business intelligence with artificial intelligence and machine learning, the possibilities are endless and exciting. With AI integrated into business and customer experiences, the hope is that every business user will be more productive and empowered and that every customer experience will be exciting, resulting in new levels of customer satisfaction and engagement and new growth opportunities for companies.

AI applications present a compelling way to implement AI in the enterprise. AI apps are different from traditional dashboards and business applications in that AI apps are designed together with domain experts to meet their specific needs for descriptive, predictive, and prescriptive insights. AI app development is accelerated using AutoML and rapid prototyping frameworks so that organizations can scale AI access across the business. The results of AI applications are tangible, such as helping patients find the right physician, reducing waste in fruit shipments, or optimizing insurance underwriting.

The future, however, is far from certain. Barriers remain, both technological and organizational. To reach this transformational future will require a new wave of innovators and leaders with the vision to find the right technology partners and create the AI applications that will drive their business and even their industries for years to come.

About the Authors

Dan Darnell is a seasoned product marketer with over 20 years of experience in leading technology companies. For the past nine years, he has been working on AI platforms and applications, including senior roles at H2O.ai, DataRobot, ParallelM, Talend, and Baynote. Before that, Dan was focused on analytics and optimization technologies at Adchemy, Interwoven, Oracle, and Siebel Systems. He holds an MBA from Carnegie Mellon University and a BS in engineering from The University of Colorado at Boulder.

Rafael Coss is Director of Technical Marketing and Developer Relations at H2O.ai in Mountain View, CA. Before joining H2O.ai, he was Director of Technical Marketing, Community, and Data Evangelist at Hortonworks. He also has served as the DataWorks Summit program cochair for three years. Before Hortonworks, he was a senior solution architect and manager of IBM's World Wide Big Data Enablement team and coauthored *Hadoop for Dummies*. At IBM, he also had roles in technical product enablement, quality engineering, and software development across multiple products and initiatives: XML database tools, federated database, and object-relational database. He holds an MS in computer science and a BS in civil engineering, both from California Polytechnic State University.

Patrick Hall is principal scientist at bnh.ai in Washington, DC, a boutique law firm focused on data analytics and AI. Patrick also serves as a visiting professor in the Department of Decision Sciences at George Washington University and as an advisor to H2O.ai.