VIEW POINT

HOW TO OPERATIONALIZE ADVANCE DATA SCIENCE AT SPEED AND SCALE FOR COVID AND BEYOND

Abstract

As agencies become more digital and connected, data boundaries blur providing agencies access to a variety of data sets in multiple formats and from multiple sources. This drives agencies to solve evolving social and population health issues, including the COVID-19 crisis, by adopting advance data-science based approaches that leverage AI/machine learning (ML) tools.

Innovation in cloud technologies, easy access to increasing computational power (CPU, GPU, Hadoop), intelligent statistical algorithms and powerful software are making it easier for agencies to adopt these advance data-science based approaches. In their survey, NASCIO found that over 55% of state IT organizations are pursuing AI initiatives and another 32% are running AI in some production operations or staging pilot projects. However, despite investing significant time, effort and resources, 87% of these AI/ML initiatives fail. Complex analytics projects must be simplified by using AI to accelerate routine tasks and to democratize data-science skills enabling organizations to decentralize analytics, making it more accessible and self-service based.

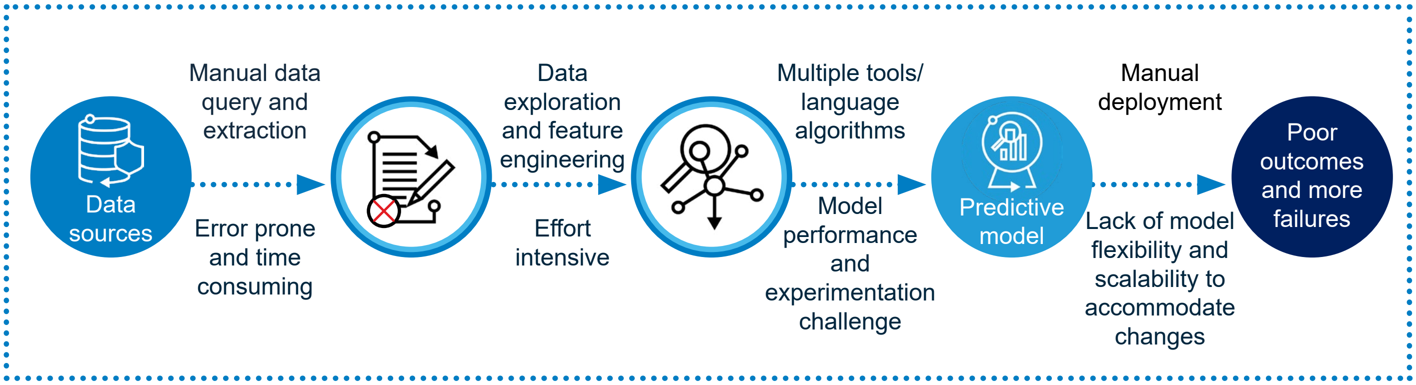
Infosys®
Public Services

The AI/ML journey and the roadblocks

Agencies are taking steps to become more data driven and analytics oriented. This involves adoption of new data science technologies like AI/ML and new approaches to foundational data management.

A typical AI/ML journey starts with identification of necessary data sources which is followed by ingestion and harmonization of relevant data, and then by the development, deployment and management of machine learning models.

In the traditional approach (manual AI/ML), agencies face roadblocks at each stage. Some of the key issues include:


Data sources
Manual data query and extraction
Error prone and time consuming
Data exploration and feature engineering
Effort intensive
Multiple tools/ language algorithms
Model performance and experimentation challenge
Predictive model
Manual deployment
Lack of model flexibility and scalability to accommodate changes
Poor outcomes and more failures

1. **Inability to onboard and process high-quality data**
   Access to clean, meaningful data representative of the problem to be solved is critical to the success of ML initiatives. Most of the existing data tend to be noisy, inconsistent, outdated, riddled with errors and difficult to process through existing tools/technologies.
2. **Complexity of data pipeline**
   Data, spread across disparate sources in different formats and protocols, must be blended and consolidated. Agencies have a hard time extracting, cleaning, standardizing, normalizing, and preparing data for predictive analytics. Data pipes grow in concert with the underlying volume of data making management of data flows from one system to another increasingly complex.
3. **Increased operational workloads**
   Increased demand for computations and complex ad-hoc analytical requests are driven by the growing volume of data. Significant time is spent on data exploration, wrangling, and feature engineering before any statistical models can be built. Most of these tasks are manual leading to an increase in the overall project's time to value.
4. **Extensive time and effort required to build and train models**
   Building ML models remains a time-consuming exercise. Not only are agencies constrained by lack of right skillets but also are straddled with legacy tools that make it difficult to define ML models, choose the right algorithms, and train, optimize and build the models needed to solve a specific business use case or validate a hypothesis.
5. **Balancing model accuracy and interpretability**
   Traditional data science projects tend to follow a 'black-box approach' to model development. This makes it difficult to understand and interpret the models or determine if they are ethical and un-biased. The result is a limited ability to regulate influencing factors and generation of imperfect insights.
6. **Operationalizing and deploying ML models at speed**
   ML models require smart deployment and continuous maintenance. They need to be regularly refreshed to ensure they perform as designed. This last mile activity, which is crucial for generating the right insights from the ML journey, is a manual and time-consuming exercise. Deployment also includes monitoring the model performance, capturing any degradation, and updating models as necessary with new data feeds. Versioning, governance and model retraining remain a big challenge as the models in production are connected with business applications and make predictions using live data.

These challenges prevent agencies from building and deploying the right models quickly. This can have serious consequences, especially if the models are being used to answer questions like the resurgence of the next COVID wave, identification of people at-risk for substance abuse or mental health issues or ordering critical PPE supplies.

Shifting the paradigm – traditional ML to automated ML for advance data science

Advancement in data science technologies can help agencies evolve from handcrafted, traditional ML to automated ML. They can build, deploy and modify AI/ML models in seconds or minutes instead of weeks. This approach can accelerate front-end data management and preparation through continuous data streaming, more seamless management of the data pipeline and continuous data quality improvement. Aggregated data can be made ready for immediate analytical use and can help generate insights faster with more confidence.

An automated data science solution enables faster adoption of ML like complex data-science modeling, enabling agencies to build more interactive ML applications for more accurate predictions. It removes all the complexities, manual effort, and chance of human error by intelligently automating the full data science model development cycle – from raw data ingestion and harmonization, data exploration, feature engineering, model training, validation, tuning and development to faster deployment and maintenance of machine learning models, and dissemination of insights.

The solution can also help build explainable and responsible AI models on the fly. These models help end-users understand and interpret how the predictions work and the reasoning behind those predictions.

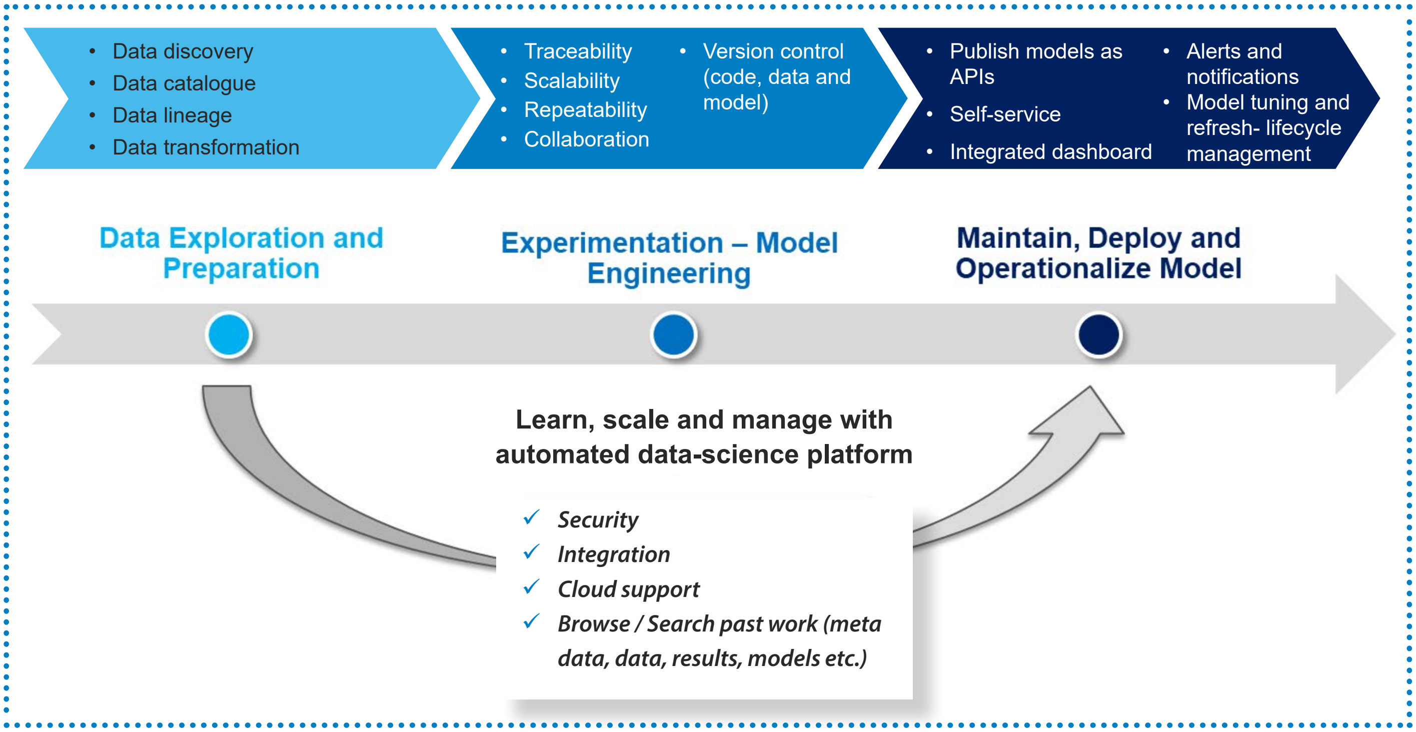
Such a solution, in addition to the benefits above, further facilitates:

- Simplified model development and validation through automated execution of several test algorithms
- Improved speed and scale in model production
- Improved model governance, maintenance and deployment
- Bulk data-handling and faster processing of data for real time analytics
- Ability to explore and compare individual models
- ML models using real-time analytics on incoming transactional data
- Reducing the cost and complexity of maintaining the data analytics platform and infrastructure

Operationalizing advance data science with an automated platform

By automating the key tasks of the predictive model development journey, an automated data-science platform can help agencies operationalize AI/ML at speed and scale. Agencies should look for these 5 key characteristics when evaluating an automated data-science platform:

1. **Enterprise scale:** Supports multiple personas and enables them to scale their analytics projects. For example, the single platform can be used by a business analyst for data exploration and preparation, a data engineer for building a data pipeline and resolving data anomalies, and a data scientist for model creation, training and deployment.


• Data discovery
• Data catalogue
• Data lineage
• Data transformation
• Traceability
• Scalability
• Repeatability
• Collaboration
• Version control (code, data and model)
• Publish models as APIs
• Self-service
• Integrated dashboard
• Alerts and notifications
• Model tuning and refresh- lifecycle management
Data Exploration and Preparation
Experimentation – Model Engineering
Maintain, Deploy and Operationalize Model
Learn, scale and manage with automated data-science platform
✓ Security
✓ Integration
✓ Cloud support
✓ Browse / Search past work (meta data, data, results, models etc.)

2. **Self-service:** Enables all users to execute their activities without any dependence on a third-person/team. It provides all the necessary tools and processes needed to support the entire AI/ML process in an automated manner without any need of programming.
3. **Unified experience:** Delivers a single window to support the entire AI/ML journey and the transparency to track every aspect of model development and operationalization.
4. **Collaboration:** Enables collaboration among stakeholders responsible for different aspects of the data-science project including model building and dissemination of insights.
5. **Democratizes data-science skillsets:** Delivers an easy-to-use, intuitive experience that enables any person with basic analytics skillsets to develop the AI/ML models they need quickly and without any specialized training.

AI-enabling different program areas with an automated data-science platform

As ML proliferates and the models grow in complexity, automated data science platforms make it easier for agencies to use ML and AI more effectively. These platforms can allow agencies to AI-enable different program areas, improving service delivery effectiveness and transforming outcomes. Here are a few use-cases that an automated data-science platform can support:

COVID

COVID Wave Prediction – Dynamic epidemiological model to predict infection projection of the next wave

Return to Work Analytics – Analyze workplace safety plans, risk scores and assess workplace infection transmission risk

Hospital Facility Care Service Analytics – Automated time series modeling on occupancy levels at hospitals, waiting times, case average, time in the ICU, etc.

HEALTH AND HUMAN SERVICES

Mental Health, Opioid Addiction and Suicide Prevention – Predict at risk population for suicide/mental health issues through analysis of early warning signs for targeted and proactive intervention

Child Welfare – Predictive modeling to discover threats through causal entity relationship analysis and determine proactive strategy to ensure child security

Fraud Detection – Identify fraudulent claims to drive investigation of types of fraud, patterns and associated provider

DMV / DOT

Traffic Management – Predict traffic volume and patterns, analyze driving trends for high risk drivers

Smart Parking Analytics – Predict number of parking spaces available at any given time

Safety Measure Analytics – Predict traffic operations and safety performance measures for compliance

SOCIAL POLICY

Enrollment Forecasting – Assess trends (increased enrollment, decreased enrollment, change in statutory benefit amount, etc.) in social programs to predict future increases in problems- medical, assistance request, etc.

Social Program Benefit – Predict future social program benefit cost increases based on economic or social conditions

Conclusion

Agencies are turning towards artificial intelligence (AI)/machine learning (ML) models to make more informed decisions across multiple program areas. However, they struggle to build and operationalize these models quickly. Various studies have found that organizations take months to develop machine learning models and that most of these models (~87%) never make it through the very complex and time-consuming process.

An automated data-science approach can help agencies automate the foundational data management activities and the AI/ML journey, enabling citizen data scientists to build, train and deploy effective machine learning models across various program areas quickly -- arming policymakers with timely insights in an environment where real-world data is changing rapidly.

From democratizing data science skillsets (i.e. making it easier for anyone to build ML models) to accelerating the model deployment lifecycle, this approach can help public sector organizations effectively use AI to improve program outcomes and citizen experience.

For more information, contact askus@infosyspublicservices.com

Infosys®
Public Services

Infosyspublicservices.com

Stay Connected