Overcoming Common AI/ML Challenges

Agencies should also consider implementing their data pipelines and machine learning with a consistent set of data governance and security.

The field of artificial intelligence (AI) focuses on creating machines that can sense, react, act, and adapt in a way that imitates human behavior. Machine learning is a subset of AI and involves enabling machines to learn from data without explicitly programming any type of rules. Instead, the machines rely on a combination of supervised, unsupervised, and reinforcement learning algorithms to detect and isolate patterns within the data which can later be used to predict future outcomes.

Nasheb Ismaily
Principal Solutions Engineer
Cloudera

The United States Public Sector is uniquely positioned to leverage both machine learning and artificial intelligence capabilities to better serve the citizen population, however what we have seen is that federal, state, and local governments must first overcome a common set of challenges in the adoption of this technology.

These challenges include ingesting data that arrives in different formats and with a diverse set of protocols, providing a holistic storage engine for both batch and streaming data, preparing data using highly parallel and distributed computation engines, and operationalizing machine learning models while providing consistent security and governance across the data lifecycle.

Meeting The Challenges

To overcome the data ingestion challenge, it's recommended that agencies leverage a single solution which can handle data arriving in batch or streaming with all types of formats including structured, semi-structured, and unstructured data. In addition, the solution should provide a low-code interface which will enable agencies to quickly onboard new data sources that can be fit into machine learning models.

For data storage, agencies should consider leveraging a highly distributed object store that can elastically scale to handle increasing data volumes. This storage engine should also provide open API's enabling connectivity from a wide range of applications. In terms of data preparation, it's recommended that agencies leverage a distributed parallel processing framework and GPU acceleration to quickly process data on a large scale.

For machine learning, agencies should leverage a solution that provides data scientists with the flexibility to develop machine learning models using the programming language and editors of their choice. This allows developers to leverage their existing skill sets to quickly train, test, develop, and operationalize machine learning models.

Agencies should also consider implementing their data pipelines and machine learning with a consistent set of data governance and security. Data governance enables agencies to track the lifecycle of the data that is used to train and develop machine learning models. With data governance, organizations will understand where the data came from, when it arrived, how it was transformed, and who was involved in its processing.

Validity and Veracity

This provides confidence in the validity and veracity of the machine learning models through traceability. With data governance, agencies can better understand patterns and changes relating to the underlying data that is used to train machine learning models. This provides a mechanism to overcome model drift, enabling developers to create automated machine learning pipelines that re-train models based on new data and confidence thresholds. In terms of enterprise grade security, agencies should leverage a single solution that provides attribute-based access controls (ABAC) across the entirety of the data lifecycle.

With ABAC, agencies can ensure that all operations used to collect, store, process, and train data for machine learning are controlled at a very fine level. Agencies should also be able to mask sensitive information based on rows, columns, and even filters within their data. Finally, all data access should be audited, ensuring only privileged users are accessing sensitive information at any given time to train machine learning models.

Cloudera is uniquely positioned to meet these requirements while providing consistent end-to-end security and governance through their Data Platform, Data Flow, and Machine Learning solutions. Example implementations of machine learning with Cloudera's technology include predictive maintenance for military vehicles where real-time sensor data is leveraged to predict equipment failure before it occurs, healthcare fraud where historical patient information is used to identify inaccurate claims, and tax fraud where the IRS is analyzing historical tax returns to identify underreported assets.

Artificial intelligence and machine learning are continuing to evolve. By leveraging Cloudera and adhering to these best practices, agencies can ensure that the ongoing innovation in these fields will continue to drive, impact, and support their missions across the public sector. ■

About The Author

Nasheb Ismaily is a principal solutions engineer and a streaming subject matter expert at Cloudera. He has over 10 years of experience designing and implementing large scale streaming analytic solutions across the public sector. Nasheb holds a dual master's degree in computer science and machine learning from Georgia Tech. He also serves as a Professor of Data Science at Regis University.

I DISCOVERED

how to prevent maintenance problems from becoming safety problems.

It's not your data. It's how you use it. Whether pushing the envelope of aerospace design or delivering vaccines years ahead of schedule, harnessing data to transform your business requires the power of artificial intelligence and machine learning to translate complex sets of information into clear and actionable insights. Cloudera's enterprise data cloud platform accelerates data analytics at every stage of the data lifecycle, with security and governance built in, to make your hybrid cloud move your business.

Learn more at **cloudera.com/datamovesyou**

#cloudera.com/publicsector

CLOUDERA
Data That Moves You