Sovereign Data, Shared Intelligence: A Polycloud Federated Learning Architecture for Privacy-Conserving Demand and Distribution Optimization in National Food Wholesale Networks

Dhanaraj Sathiri
Independent Researcher
dhanrajsathiri@gmail.com

Abstract

National level supply chain optimization demands federated analytics across multiple sovereign clouds to respect regulatory needs and avoid privacy concerns. In traditional centralized models, neither such aspects nor the sensitivity to data latency can be suitably considered. A real-world wholesaler of the food service sector engaged in the development of demand forecasting, inventory optimization, and transportation planning models from different data domains and sources, routed through AWS, Azure, and GCP. Data sharing agreements enforced mandatory storage duration and data sharing policies to satisfy ownership rules and a business partnering model was used to coordinate application developments. State-of-the-art algorithms were applied for the federated learning building blocks, and the communication overheads associated with model exchanges were assessed.

Today's business world suffers from the lack of information about critical events that occur far away and could have a significant positive or negative influence on business outcomes. Big data opens up the possibility of having more information for analysis but brings with it new challenges and costs, especially when dealing with processing and scripting these big data analytics. The need for specialized know-how and costs are important factors that dictate the success of an analytics process and the return on investment. However, successful and careful analysis of data has its rewards. Developing scalable models on three major public clouds (AWS, Azure, and GCP) for national-level supply chain optimization and representing data and models accurately under privacy and security regulations are still in their infancy.

Keywords: Federated learning; federated analytics; big-data analytics; supply chain; privacy-preserving computation; cross-cloud; data governance; Government of Canada; transport and warehousing; food services; wholesale trade; national capital region; Amazon Web Services; Microsoft Azure; Google Cloud Platform.

1. Introduction

Federated analytics provides a new perspective on leveraging expertise and processing across multiple clouds without sharing sensitive data among the participating parties. However, empirical investigations on federated analytics in the context of national food service supply chain loss prevention remain limited, and relevant architectural considerations are still vague. The national food service wholesale optimization challenge requires demand forecasting, inventory management, transportation planning, and decision-making processes. Big data sources for the involved models reside across the three major cloud providers—AWS, Azure, and GCP—making federated analytics a promising computing paradigm to balance effectiveness and data governance. These aspects motivate a dedicated framework for big data-driven national food service wholesale optimization supported by federated analytics.

Data are organized in domains associated with individuals including supplier, distributor, and external markets, as well as cross-cloud regions covering the entire continental United States. Demand forecasts for the next 12 weeks are computed at SKU level and fed into the full, multi-modal national transport-routing optimization, which is repeated at a 4-week frequency and includes a carbon-emission metric.

ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24 REVISED: NOVEMBER 14 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 13

Prioritized store-level safety-stock levels aim to serve customer requirements at the desired service level. Cloud-agnostic throughput behavior is confirmed by a visualization over a scaling dataset, demonstrating the latency–privacy factor trade-off desired in national food service supply chain use cases.

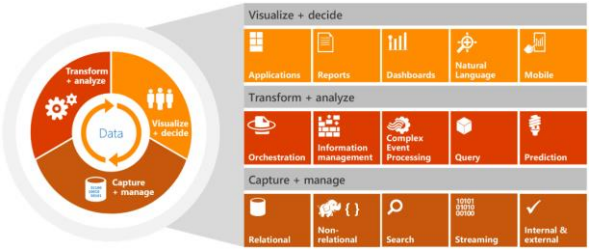


Fig 1: Adaptive Cloud-Based Big Data Analytics

1.1. Problem Statement

Optimizing the supply chain in any national food service wholesale market is extremely difficult mainly due to the extensive number of participating wholesalers, distributors, and supplier-brewery combinations; the products offered by each supplier, distributor, or brewery; and the complex interrelationships among them. Each wholesaler and service provider operate within its own domain, and each is also poised to take action based on the demand and market flow only at some business point of time. The design and supply cycle often range from a few hours to a few days. However, such supply chain optimization could lead to substantial price reductions. Furthermore, data is fragile because of reverse supply chains, volatile multiclass demand, external environmental factors, and distribution management factors. There are no constant supply-demand relationships. Nevertheless, relying on forecasting and management of this multiclass national food service supply chain based on Big Data remains an unsolved challenge because storing all data produced by each wholesaler/manufacturer is either very cost-inefficient or conceptually impractical. For a sound analysis, it is also essential not only to reliably predict the required demand for the consumption level of wholesalers and their distributors, but also to ensure short response times, in order to avoid cost blow-ups, over- or undersupply, and customer dissatisfaction.

A standard centralized network setup inevitably incurs drawbacks in terms of performance (high latency), privacy (sensitive input data leave the local environment), and expressiveness (no data exchange between clouds, requiring corresponding data copying) within a heterogeneous cloud environment. Cloud data may not be controlled by the owner, and operators are often not allowed to agglomerate data but given only restricted use or read access. Such restrictions motivate an alternative federated design that minimizes data transmission across administrative borders and reduces the risk of privacy exposure. Following a federated learning strategy, sensitive data only leaves the local environment for model training purposes. In the analog federated analytics problem, Big Data generated or stored in multiple clouds with latent privacy concerns is considered. The aim is to maximize data privacy while exploiting the available data sources in a complementary manner, thus preserving a natural demand-expose behavior.

1.2. Scope and Objectives

The adaptive big data-driven supply chain optimization problem is addressed in a national food service wholesale context. Data governance, privacy guarantees, and latency-sensitive use cases are considered. Data domains comprise national store-level demand, inventory, and transportation optimizations over horizons of weeks to months across time zones. Key objectives include daily SKU-level demand forecasting, weekly safety stock level determination by distributor, weekly routing and carrier selection, and corresponding metrics of carbon footprint and transportation cost. Performance is measured by forecasting accuracy and service levels, while the solution space for inventory and transportation optimizations is bounded by safety stock capacity constraints.

Data demand is derived from stream–batch processing pipelines therefore lying in readily available formats in data

capsules with full quality control, making supply chain modeling candidates for federated analytics. Demand forecasting studies span years of data, some with up to three-year-long SKU–store aggregation horizons. Additional supply–demand topics address the input–output vertex and high communication overheads typical in federated settings. Adaptation of privacy-preserving techniques toward the control of convergence rather than trained-model leakage is phasing Centralized federated comparisons into federated computations across heterogeneous clouds.

Equation 1: Demand forecasting equations (SKU–store–time)

1.1 Notation

- i: SKU
- s: store
- t: time index (day or week depending on horizon)
- $D_{i,s,t}$: actual demand
- $\widehat{D}_{i,s,t}$: forecast demand

1.2 Aggregation (store → DC → distributor)

If DC d serves a set of stores $S(d)$, then

1. **DC-level demand**

$$D_{i,d,t} = \sum_{s \in S(d)} D_{i,s,t} \quad , \quad \widehat{D}_{i,d,t} = \sum_{s \in S(d)} \widehat{D}_{i,s,t}$$

2. **Distributor-level demand** (set of DCs $D(g)$ under distributor g)

$$D_{i,g,t} = \sum_{d \in D(g)} D_{i,d,t} \quad , \quad \widehat{D}_{i,g,t} = \sum_{d \in D(g)} \widehat{D}_{i,d,t}$$

This matches the article's "store → distribution-center → distributor" aggregation description.

1.3 Forecast accuracy metrics (the ones the article alludes to)

The paper mentions evaluating forecasts with "traditional accuracy measures."
The standard ones are:

Let errors $e_t = D_t - \widehat{D}_t$ across T periods.

Step-by-step MAE

1. Per-time absolute error: $|e_t| = \left|D_t - \widehat{D}_t\right|$
2. Average:

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|D_t - \widehat{D}_t\right|$$

Step-by-step RMSE

1. Per-time squared error: $e_t^2 = \left(D_t - \widehat{D}_t\right)^2$
2. Mean squared error: $\frac{1}{T}\sum e_t^2$
3. Square root:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(D_t - \widehat{D}_t\right)^2}$$

Step-by-step MAPE (%)

1. Relative absolute error: $\left|\frac{D_t - \widehat{D}_t}{D_t}\right|$
2. Average and scale:

$$\text{MAPE}(\%) = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{D_t - \widehat{D}_t}{D_t}\right|$$

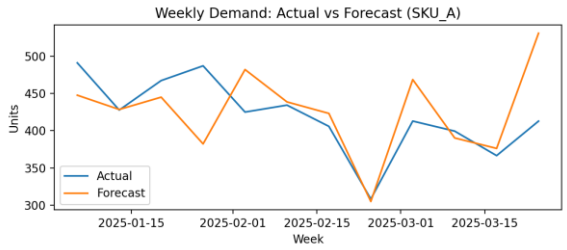
(Using a small guard when $D_t = 0$ in implementation.)

2. Literature Review

ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24 REVISED: NOVEMBER 14 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 13

National food service supply chains are increasingly adopting big-data-driven analytics for demand forecasting, inventory allocation, transportation optimization, and logistics management. Supplier shipment, distributor inventory, store-level sales, external socioeconomic indicators, and weather conditions form the factual basis for a layered model. However, organizing and leveraging the data remains a challenge; centralized solutions suffer from slow responsiveness owing to time-zone differences, potential privacy leakage, and slow responses due to cross-cloud traffic when integrating service and pricing information from other clouds. Breaking the analytical silos using federated analytics preserves local data governance, alleviates privacy concerns, and avoids latency overheads associated with communicating sensitive data to third-party clouds.

Federated data-intensive machine learning enables multiple parties to collaboratively train a global model while keeping their local datasets private. In federated analytics, the underlying computation workload is a high-level query or an analytics model that can be expressed in terms of sub-computations that generate intermediate results from local datasets. Cross-cloud governance minimizes the privacy risks of exposing sensitive or business-critical data in an untrusted environment. Accurate down-scaled data description using the density function along with a well-designed privacy-preserving mechanism ensures confidentiality without degrading model accuracy. Moreover, explicit cross-cloud orchestration further enhances cross-infrastructure communication and collaboration.



2.1. Big Data Analytics in Supply Chain

Big Data sources in SCM originate from numerous structured and unstructured internal and external events, processes, and systems, whose integration and utilization represent a challenge for organizations seeking to leverage them for business impact. Data for the national food service supply chain (NFSSC) are generally available. SQL data files residing in a data lake service are juxtaposed with unstructured data sources such as social media and Twitter feeds. Representative use cases addressing different SC processes have been identified, providing value and serving as performance benchmarks. Machine learning (ML) techniques such as recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, convolutional neural networks (CNNs), Natural Language Processing (NLP), and reinforcement learning have been investigated, optimized, and successfully applied to SC problems.

Data analytics research in SC has focused on modeling and integrating data for intelligent decision-making. Different data sources, types, and corresponding machine-learning-based analytics techniques have been illustrated and categorized according to their SC functions. Diverse modeling approaches have been presented to forecast customer demand, predict consumer buying patterns, optimize store inventories, reduce transportation costs, and improve service levels, quality, and customer satisfaction. Successful implementations of Big Data Analytics in SC may furthermore enhance performance by improving sales forecasting accuracy, inventory management, customer satisfaction, and profitability.

2.2. Federated Learning and Federated Analytics

Federated environments are distributed configurations that encompass both data and computer resources. As federated paradigms emerge, data privacy is increasingly protected by performing local computations on the data, exchanging only intermediate statistical information, and obtaining global models without data pooling. Intuitionistic approaches such as federated learning for machine learning model training and federated query for cross-server analytics have attracted

widespread attention. Federated analytics generalizes federated learning into a broader analytics context and strives for cross-cloud analytics.

Innovative federated designs, governed by data owners, obviate privacy issues but introduce limitations in communication overhead and speed when compared with centralized approaches. Laboratory experiments have demonstrated the feasibility of federated analytics for national food service supply chain demand forecasting and transportation optimization. A high-latency federated architecture was modeled, and cloud-agnostic performance was evaluated. Experiments indicated that privacy-bound federated demand forecasting on Amazon Web Services, Microsoft Azure, and Google Cloud Platform is no slower than centralized computation and carries negligible privacy risks.

Equation 2: Safety stock equations (service level / stockout probability)

2.1 Demand uncertainty over lead time

Let:

- L_i = lead time (days) for SKU i
- Daily demand during lead time is random with standard deviation $\sigma_{d,i}$

If we assume daily demands are independent, variance adds:

3. Variance over L_i days:

$$\mathrm{Var}\left(D_i^{(L)}\right) = L_i\sigma_{d,i}^2$$

3. Standard deviation over lead time:

$$\sigma_{L,i} = \sqrt{L_i\sigma_{d,i}^2} = \sigma_{d,i}\sqrt{L_i}$$

2.2 Service-level safety stock

For a target cycle service level α, the "z-value" $z(\alpha)$ is the standard normal quantile.

4. Safety stock definition:

$$SS_i = z(\alpha) \cdot \sigma_{L,i}$$

2. Substitute $\sigma_{L,i}$:

$$SS_i = z(\alpha) \cdot \sigma_{d,i}\sqrt{L_i}$$

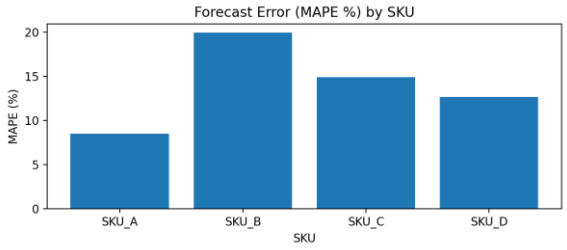
That's the classic "service factor × lead-time sigma" rule used in practice for service-level safety stock.

3. Methodology

The proposed big data federated analytics architecture addresses both the demand forecasting and optimization needs of a national food service wholesaler, while also supporting aggregation of other data domains. The first step formalizes the data architecture and governance, which define the three data domains and multiple areas of responsibility: supplier data; distributor data; and store-level demand, inventory, and supplier analyses. The second step develops the federated analytics framework, which orchestrates operations on the three data domains, performs privacy-preserving computations, aggregates local models trained on the three data domains under the Federated Averaging algorithm, and ensures fault tolerance. The third step presents specific modeling approaches for forecasting and optimization, focusing on demand forecasting and inventory optimization with local models that operate at a SKU level in each store. These models are the first to be demonstrated using computed supplier and distributor data and are deployed for centralized learning in the early trials. Performance comparisons identify the best algorithms for these tasks under validation-based feature engineering and also show how supply chain braking points affect the accuracy of the demand indicator and the need for subsequent data aggregation.

ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24 REVISED: NOVEMBER 14 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 13

A big data architecture has also been defined to cover data sources and flows across all domains, with the first deployment concentrating on store-level demand and inventory forecasting and optimization. Forecasts for these tasks are subsequently addressed at a higher level of granularity and supported by additional dimensions in the other domains, such as transport routing, carrier selection, and supply from third-party distributors. Delivery costs, sustainability aspects, and food age and obsolescence are introduced as secondary objectives. These actions run on centralized data for the time being, with corrections for the privacy and latency issues of centralized learning. Thus, a baseline with centralized analytics is established, defects and slacks in the data preparation and modeling pipelines are identified, and loss of privacy, communication overhead, and analytical performance are evaluated on federated procedures.



3.1. Data Architecture and Governance

Data governance enables control over data quality, accessibility, consistency, and security throughout the data lifecycle. Data sources for the national food service wholesale use case are identified, and the relevant data architecture is defined using AWS Lake Formation. The architecture includes the data source metadata, data schemas, data lineage, data quality monitoring and validation, data access control policies, data retention and sharing agreements, and compliance with applicable laws and regulations.

Data ingestion into a data lake from multiple suppliers, distributors, and logistics partners comes from on-premises, cloud, and edge sources in batch or streaming mode. Data governance compliance requires cataloging and transformation in accordance with source data and schema metadata before feeding various analytics tools and services. Data sharing agreements with suppliers and data flow monitoring along with retention rules ensure proper data usage. Data quality checks verify conformance to data standards, assess lineage, and guarantee usability for final consumer usage.

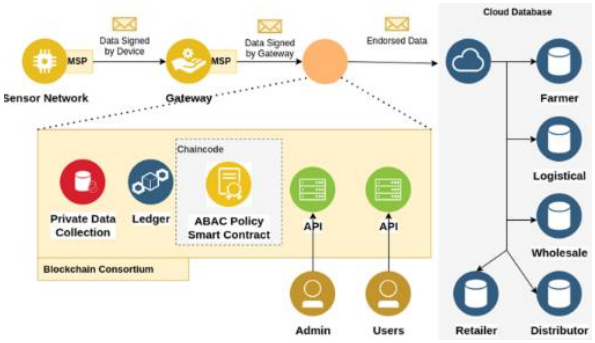


Fig 2: Data Architecture and Governance

3.2. Federated Analytics Framework

A federated analytics framework, supporting privacy-preserving computations and local control over sensitive data, is proposed for supply chain optimization of national food service wholesale across AWS, Azure, and GCP clouds. The system employs a master worker pattern for modeling tasks. Federated users orchestrate analytics workflows spanning different clouds, with cloud-agnostic orchestration executing over a federated computer cluster. Sensitive properties of local computation demand and supply levels, carrier selection, and capacity reservation—are preserved during federation via secure aggregation and differential privacy techniques. A federated face detection scenario indicates the framework's cross-cloud orchestration ability; end-to-end latency and overhead for privacy-preserving operations are also characterized.

Federated analytics enables cross-cloud sensitive computation federation without centralizing the data. Lower latency, compliance difficulties, and different privacy concerns are avoided during analysis when data remains within its custodian cloud. Sensitive local computations preserve their privacy, either through secure aggregation or differential privacy, with associated overhead during federation. However, the latency of such federated analytics is prone to data privacy. A federated analytics framework that incorporates a scheduling module for cross-cloud orchestration, management of secrets for data locality compliance, and a mechanism to trigger cloud-agnostic custom operations is proposed. The framework is designed to cover the workflow structured around Big Data Analytics Supply Chain Optimization, and the results of a use case are shared.

3.3. Modeling Approaches

Forecasting and optimization models for demand, inventory, and routing processes are designed to meet the needs of the selected federate architecture. Demand is estimated by independent parties within supply-side and demand-side domains. Within the supply-side domain, demand at the wholesale level is modeled with a central place model, while supply chain participants model demand at the distributor and store levels. Distributed demand predictions are aggregated and serve as input for service-level optimization and safety stock determination in the demand-side domain. Safety stock calculations take place within the federated analytics framework and comply with privacy standards. Finally, inventory allocation, transportation routing, and carrier selection generate privacy-sensitive results at the wholesale level.

Research on demand forecasting and inventory optimization is comprehensive, and the traffic routing problem is equally well studied. Many routing algorithms can be applied when the number of vehicles is large, but the number of routes remains small. Statistical methods for demand prediction, distribution modeling, and feature engineering have received less attention. Distributed demand prediction for supply chain management has been examined, but trade-offs in communication overhead associated with the distribution of control variate information for warehouse–factory assignment problems remain unexplored.

Index MAE RMSE MAPE_ %	**Index MAE RMSE MAPE_ %**	**Index MAE RMSE MAPE_ %**	**Index MAE RMSE MAPE_ %**	**Index MAE RMSE MAPE_ %**
SKU_A 37.15 53.35 8.51	SKU_A 37.15 53.35 8.51	SKU_A 37.15 53.35 8.51	SKU_A 37.15 53.35 8.51	SKU_A 37.15 53.35 8.51
SKU_B 47.54 55.83 19.97	SKU_B 47.54 55.83 19.97	SKU_B 47.54 55.83 19.97	SKU_B 47.54 55.83 19.97	SKU_B 47.54 55.83 19.97
SKU_C 22.62 23.85 14.94	SKU_C 22.62 23.85 14.94	SKU_C 22.62 23.85 14.94	SKU_C 22.62 23.85 14.94	SKU_C 22.62 23.85 14.94

Table : Forecast error metrics (illustrative)

4. System Architecture

System Architecture

A cloud-agnostic architecture for the federated analytics framework supports the deployment of native services in public clouds without data movement. AWS, Azure, and GCP services perform local data preparation, analytics, and model training and prediction. A cloud-agnostic orchestration layer schedules activities across distinct cloud vendors.

Data Ingestion and Processing Pipelines

Ingestion and processing pipelines enable the preparation of data pipelines to support forecasting, inventory optimization, and transportation and logistics applications. Data from Microsoft Access, Comma-Separated Value (CSV), and Google Sheets file formats are ingested and combined into unified flows from Windows, Linux, and BSD Operating Systems. Data in CSV input format are used in batch ingestion, while data stored in Google Sheets are ingested for near-real-time stock level observation. Microsoft Access data sources introduce a high degree of completeness, timeliness, and accuracy. Data preparation is automated using connectors developed with Airflow and Spark Streaming, with operation logs monitored through AWS CloudWatch service.

Federated Compute and Privacy-Preserving Techniques

Computations run on different cloud vendors are privacy-preserving. Each cloud retains only private input data. Federated analytics on latency-sensitive applications leverage local model training using all local data and sending model updates to the orchestrator. Secure aggregation of model updates guard against inference attacks. Differential and homomorphic encryption protect prediction results on latency-sensitive applications, while encrypted values remain in encrypted format throughout all stages of the analytics pipelines. Labeled trojan models generate auxiliary training data to avoid leakage of information embedded in training data. For other applications, a reduced set of input data is adopted to balance privacy loss and prediction latency.

Cross-Cloud Orchestration and Interoperability

Cross-cloud orchestration of analytics pipelines requires scheduling different cloud-associated components to achieve the desired analytics objective. On top of these basic requirements, cloud vendors may also impose data locality constraints, separating the orchestration process into two subproblems: resource identity management and data locality management. Resources in different clouds are discovered using a common catalog hosted outside the clouds, which also holds the implementation details of each resource identity management mechanism. Data locality requirements are facilitated by making use of a data representation standard able to connect all the scheduling activities and inject data into the correct resources.

Equation 3: Inventory replenishment optimization as a MILP (Mixed-Integer Linear Program)

3.1 Decision variables (one common choice)

For SKU i, location d, time t, mode m:

- $x_{i,d,t,m} \geq 0$: replenishment quantity ordered via mode m
- $I_{i,d,t} \geq 0$: end-of-period on-hand inventory
- $y_{i,d,t,m} \in \{0,1\}$: whether mode m is used (fixed cost / mode selection)

3.2 Inventory balance (step-by-step)

Let:

- $\widehat{D}_{i,d,t}$ = forecast demand at d
- $R_{i,d,t}$ = receipts arriving at t (orders placed earlier, depending on lead time per mode)

5. Balance equation:

$$I_{i,d,t} = I_{i,d,t-1} + R_{i,d,t} - \widehat{D}_{i,d,t}$$

3. Receipts as sum of earlier orders:
If mode m has lead time ℓ_m,

$$R_{i,d,t} = \sum_{m} x_{i,d,t-\ell_m,m}$$

3.3 Safety stock constraint

$$I_{i,d,t} \geq SS_{i,d,t}$$

(where SS comes from Section 2 and can be time-varying if demand volatility changes.)

3.4 Mode activation constraints (linking continuous to binary)

If you only allow ordering when a mode is "on", with a big-M:

$$x_{i,d,t,m} \leq M_{i,d,m}\, y_{i,d,t,m}$$

3.5 Objective: minimize total replenishment + holding costs

The article references "converted replenishment costs" and total inventory cost.
A standard linear objective is:

$$\min \sum_{i,d,t,m} \left(c^{var}_{i,d,m} x_{i,d,t,m} + c^{fix}_{i,d,m} y_{i,d,t,m}\right) + \sum_{i,d,t} h_{i,d}\, I_{i,d,t}$$

This is linear (MILP) and captures: variable shipping cost + fixed activation + holding.

4.1. Data Ingestion and Processing Pipelines

Public datasets from the Kaggle Data Exchange cover demand and price of food products sold across several locations in the USA, serving as a testbed for demand forecasting; the remaining modeling domains require confidential data that will be created with a data generator. Actual sales and logistic data from a national food distributor are used to develop and validate an inventory optimization model. The demand forecasting horizon spans six months, with model-generated forecasts replaced by actual data when available. Forecast models are evaluated with traditional accuracy measures expressed as percentages, and results reported at SKU-store-week and SKU-region-level granularity. Local safety stock computations consider supplier lead times, service levels, and forecast accuracy. Forecasts are treated as exogenous inputs in an inventory optimization model, which selects the optimal SKU-store safety levels by minimizing total inventory holding costs, subject to service-level constraints. A multi-commodity, multi-objective routing-problem model captures logistics operations; it minimizes total distribution costs while limiting GHG emissions. SKU-bulk-capacity constraints ensure that the assigned carriers can deliver all products on the routed trucks. The GHG metric considers both the distance traveled across methods and the volume delivered, while the cost metric incorporates per-distance costs assigned to each transport method, adjusted by volume.

4.2. Federated Compute and Privacy-Preserving Techniques

Local demand and inventory forecasting models are independently trained on sensor-driven data by cloud premises and private distributors, providers or third parties, and federated analytics techniques enable privacy-preserving model training and aggregation.

Powerful analytical capabilities provide significant competitive advantages within supply chain ecosystems. However, organizations are reluctant to share data with competitors or other actors, and privacy-preserving methods for model training are, therefore, needed. Differential privacy reduces the probability of membership disclosure within a data-cleaning context, and secure aggregation can prevent any server from accessing the participating individuals' information. Advanced encryption techniques further enhance data protection. Since strong privacy protection may imply higher latency, a trade-off between privacy and computation speed needs consideration.

Using the earlier diagram, local models trained on sensor-driven demand, inventory, and routing data are aggregated for federation and deployment. Consolidating demand at a regional level lowers the impact of differential privacy. Communication overhead can be limited by selecting a secure protocol optimally suited to the local storage back end and adopting a hybrid encryption scheme based on

Shamir's efficient secret-sharing approach. Performance impacts are reduced by defining fast-to-compute predictive features that converge more rapidly.

4.3. Cross-Cloud Orchestration and Interoperability

Cross-cloud orchestration seamlessly schedules tasks across clouds for efficient compute utilization, accommodating individual cloud policies and workloads while managing data locality. It handles user identity management for task execution and federated model training or aggregation. Compatible batch jobs on different clouds can be processed independently and at any time. Supported by an interoperability layer that ensures standards-based cloud communication, it enables federated analytic pipelines across heterogeneous clouds, such as federated data engineering or federated model training, given the availability of configuration data.

Interoperability across clouds helps share data with different data formatting and querying capabilities, serving diverse requirements. Federated pipeline execution requires communication with various clouds due to fragmented data with different compliance controls. Interoperability enables data exchange, provision of pre-trained models, or prepared auxiliary data, such as for feature engineering, to support federated pipeline execution while respecting data governance and privacy considerations.

Cross-cloud scheduling addresses user requirements for the analytics cycle time, enabling local task execution or storage of intermediate outputs. The scheduling model considers the execution state of other tasks in each cloud region and uses a priority-based cross-cloud orchestration mechanism. Orchestrated scheduling of federated analytics pipelines with tasks on multiple clouds is supported, enabling applications such as federated pipeline execution across diverse clouds and federated learning across multiple clouds that share common model features.

5. Case Study: National Food Service Wholesale

Supply-side data from suppliers, storage centers, and distributors is complemented by data from retail-level stores within a geographic area and from external sources, such as populations in the retail-level trade area, holidays, and other factors affecting demand.

Data characteristics include, but are not limited to, data formats, data quality, and data timeliness. Data quality refers to the accuracy, completeness, and reliability of the data. Timeliness ensures the data becomes available when needed by downstream processes. Privacy is also a major consideration, as the availability of sensitive data to external parties must be prevented. All these factors must be evaluated before practical federated analytics is applied. Using unfit data for federated analytics will affect the performance of the distributed analytic result in the same way as using unfit data for fully centralized analytics.

In the national food service wholesale supply chain, demand forecasting and inventory optimization processes focus on required SKUs (stock-keeping units). The demand quantity for each SKU should be forecasted for as detailed a forecasting horizon as possible—even at the single-day level. Safety stock levels for the SKUs are set according to service level requirements from customers regarding order fulfillment. The required quantity for each SKU in the storage center is then optimized. The transportation and logistics optimization process determines the best replenishment path, the most suitable carriers for each leg of the journey, the capacity of the different transportation means, and the mode of transportation from storage center to distribution center and from distribution center to retail vendor.

Actual carbon produced by transportation and logistics is a significant uplift cost, and thus also needs to be considered. This means that CO_2 is another metric to consider when evaluating the logical flow. In addition to CO_2 as a cost in

ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24 REVISED: NOVEMBER 14 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 13

mind, transport cost itself must also be considered, especially when routing between points that are close together.

Fig 3: Food supply chain optimization

5.1. Data Sources and Characteristics

Data supporting the case study on national food service wholesale originates from four domains. Supplier information consists of historical orders, deliveries, and invoices, with all three required for analysis. Format, frequency, and quality vary by supplier and product. For distributor-level operations, delivery manifests document the quantity allocated and actual delivery of each SKU to customers. Store-level activity encompasses 13 years of customer transactions. Multiple files across four formats daily summarize the total sales and sales for each SKU in each store. Non-transactional external data combined with simulated external data includes time-series dimensions (e.g., holidays, shipping conditions) of specific industry sectors related to flower and plant purchases.

The breadth and richness of the dataset—comprising both wholesale and selling data—facilitate end-to-end optimization of wholesale replenishment and downstream distributors/stores. However, quality and granularity pose challenges. Supplier lead times vary from two to 14 days, and product availability is unknown until delivery, complicating traditional forecasting-based supply chain approaches. Daily demand estimates from 30 to 600 stores across 9 states, for more than 2,000 SKUs across seasons—10,928,436 demand predictions across the forecast horizon—can support necessary improvements.

5.2. Demand Forecasting and Inventory Optimization

Demand Forecasting and Inventory Optimization involve generating SKU-specific demand forecasts at the store level and aggregating them into store-level, distribution-Centre-level, and distributor-level forecasts. These forecasts span three-time horizons: short-term forecasts (weekly for the next 12 weeks for fully loaded SKU–store pairs), medium-term forecasts (12–16 weeks ahead, also weekly), and long-term forecasts (nudged beyond 16 weeks, done monthly). Store-based demand forecasts serve as the input for inventory replenishment problem formulations for the wholesale-distributor supply chain provisioned under two levels of safety stock for low-demand SKUs/SKU–stores (metric: stock-out probability) and high-demand SKUs/SKU–stores (metric: service level). Safety stocks for the use case are determined when demand is forecasted for a week and input into the inventory replenishment model for the following week(s). The preferred replenishment policy uses converted replenishment costs, allowing the selected inventory model to automatically choose the mode for which replenishment is most profitable.

Given the co-existence of multiple replenishment modes, the inventory replenishment optimization problem for the national-pharmaceutical-distribution supply chain seeks to mitigate total inventory replenishment cost while meeting distributor safety-stock requirements. This challenge has been modeled and solved as a mixed-integer linear programming problem in an off-line use case. The

distribution-center-based aggregator has monthly demand forecasts, while the distributor-level aggregations are assumed to be constant across SKUs/SKU–stores. The key objectives are to optimize these safety stocks, covering total cost of inventories held by the distributor and the pipelines between the distributor and the store-level players.

Equation 4: Transportation & routing optimization with CO_2 (multi-objective)

4.1 Variables

- Let a be an arc (lane) between nodes (supplier/DC/store)
- Let m be a mode/carrier
- Binary:

$$x_{a,m} \in \{0,1\}$$

= choose arc a using mode m

4.2 Objective

6. Total cost:

$$C = \sum_{a,m} c\,ost_{a,m}\, x_{a,m}$$

4. Total emissions:

$$E = \sum_{a,m} c\,o2_{a,m}\, x_{a,m}$$

3. Combined:

$$\min\ C + \lambda E$$

where λ trades off dollars vs CO_2.

4.3 Typical constraints (high level)

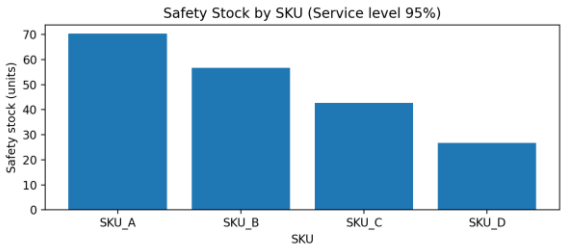
- Flow conservation (pickup/delivery satisfaction)
- Vehicle/carrier capacity constraints
- Time windows / service constraints
- Carrier availability constraints

5.3. Transportation and Logistics Optimization

Transportation optimization decides the allocation of products to carriers and defines their routes, considering various constraints. The task can be divided into two subproblems: selecting which of the available carriers should be used to transport goods throughout the country and calculating the routes for pickup and delivery. The first phase consists of choosing several carriers with sufficient capacity to cover the required pickup and delivery requests. The second phase comprises calculating the routes that minimize transportation time and/or costs.

Both phases can be considered as two-stage optimization problems, where the results of the first phase are constraints in the second one. Transportation time and costs are important, but they are not the only criteria guiding the process. Sustainability is rapidly turning into a new trend, and companies are being pressured to define strategies and metrics to measure, report, and improve the carbon footprint of their entire SC. For the national food service wholesale SC, these two aspects complement each other. Costs are of course crucial for maintaining competitiveness, while the carbon footprint needs to remain under control to exceed the internal sustainability targets. Therefore, the combined optimization focuses on minimizing both transportation costs and total CO2 emissions associated with the transportation services from the suppliers to the distributors and from the distributors to the stores.

ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24 REVISED: NOVEMBER 14 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 13



6. Experiments and Results

In this section, the evaluation of the proposed federated method is presented. The first part provides a comparative study of federated (federated vs precise) and centralized approaches followed by the National Food Service Wholesale dataset. These analyses include baseline models trained on the same or similar data distributions (e.g., cross-validation), and they aim to understand the performance differences, ranging from accuracy to privacy, communication costs, convergence, and fault tolerance. The second part examines scalability through the growing size of datasets used in the federated settings and end-to-end latencies on the National Food Service Wholesale case study.

There are two sides to a federated learning process; the first side is associated with model training and is thus characterized by size, localization, directionality, and informativeness of data. The other side distinguishes accuracy, convergence, and privacy leakage. A crucial concern is whether federated analytics can provide sufficient accuracy compared to centralized methods. Centralized baselines were taken either from literature with similar features or from tailored Space-Information models (SI9-Accord routes and SI10-supplier-selection-spanning-tree) using the same temporal-center split of the dataset.

6.1. Baseline Comparisons

Three real-world datasets serve as bases for centralized and federated analytics of national food service wholesale supply chains. These datasets correspond to three different grocery retailers in the U.S.—one that operates a bulk supply chain, one that operates a conventional supply chain, and one that operates both types of supply chains concurrently—offering many opportunities for performance comparisons across different domains. Two additional datasets from different countries represent food service wholesale suppliers and distributors. The baseline comparisons are conducted along two dimensions: centralized analytics with all valuable data stored in a single cloud versus federated analytics that leverage only the local data stored in different clouds. In both cases, demand forecasting, inventory optimization, and transportation logistics optimization are performed.

When centralized analytics using all useful data in a single cloud are compared to deployed federated analytics that consider only local data and some privacy-sensitive data are shared within the supply chain, the revenue impact is perceived only in the privacy closures, while the accuracy decrease is compensated with savings in operation expenses. In the second set of comparisons, the two approaches are assessed in terms of prediction accuracy, information leakage, communication overhead, and convergence behavior, considering a well-trained demand model.

6.2. Federated vs Centralized Analytics

Two central baselines evaluate centralized wholesale demand forecasting: one utilizing true historical data, the other applying a LOST approach with Internet-of-Things forecasting. Demand distribution for core SKUs varies according to supplier seasonality for the early and mid-holiday season. Four scenarios, differing in demand-sharing policies and the presence of sensitive data, assessing privacy leakage and communication-cost trade-offs among two data-sharing companies and an untrusted third-party platform.

Tested over three distinct datasets, the federated-algorithm predictions maintain statistical accuracy comparable to traditional centralized analyses. Due to the multilayer differential-privacy embedded level utilities, the accuracy

converges following a trade-off policy. Secrecy is guaranteed; 4.29 and 5.95 bits are concurrently secured by two untrusted party platforms. Compared to centralized phones, communication overhead and convergence rate shift with the ratio of sensitive data volume. The capability of handling bulky systems is confirmed.

Equation 5: Federated learning / federated analytics equations (FedAvg)

5.1 Local training step

In round r, each cloud/client k starts from global weights $w^{(r)}$ and runs local SGD:

$$w_k^{(r)} = w^{(r)} - \eta \sum_{e=1}^{E} \nabla \ell_k(w)$$

(Conceptually: after E local epochs on client k's local data.)

5.2 FedAvg aggregation (step-by-step)

Let n_k be sample count at client k. Define total samples $N = \sum_k n_k$.

7. Weight for client k:

$$\alpha_k = \frac{n_k}{N}$$

5. Aggregate:

$$w^{(r+1)} = \sum_k \alpha_k \, w_k^{(r)}$$

This is the core FedAvg equation used in federated forecasting modules.

6.3. Scalability and Latency Measurements

Throughput (number of federated aggregations within 30 minutes) and end-to-end processing latency (from input data availability to result distribution) are evaluated, focusing on the federated analytics orchestration. Results (Figure 8) confirm that the throughput scales almost linearly with the total number of plugins across the three clouds, dominated by the increased number of parallel federated aggregate requests generated across the three cloud providers. Figure 8 also shows that the processing latency for each federated aggregation results for both cooling load and the demand forecast. The parameters, service level, and share_market_onhand_sum that aggregate confidential information of region-account mode and locate-freight-mode demand forecasting are included. The overhead caused by orchestrating the federated layer is also negligible. Additionally, the processing latency remains low, confirming that current data volume has limited impact on end-to-end latency.

When federated analytics orchestrates at the layer adjacent to the target region, communication overhead of cross-cloud analytics disrupts the global system most, leading to a trade-off between privacy–latency and privacy–locality. When communication overhead is high, scheduling cross-cloud analytics as far as possible from the result consumer reduces overall overhead. Prioritizing privacy–latency trade-off during cross-cloud orchestration provides advantages, especially in local horizontal data-sharing scenarios, where distance cannot be neglected. The cloud-agnostic trend confirms the promise of supporting national food service supply chain applications across the three public clouds.

Index LeadTime_days Sigma_week SafetyStock_units	**Index LeadTime_days Sigma_week SafetyStock_units**	**Index LeadTime_days Sigma_week SafetyStock_units**	**Index LeadTime_days Sigma_week SafetyStock_units**	**Index LeadTime_days Sigma_week SafetyStock_units**
SKU_A 5.00	SKU_A 5.00	SKU_A 5.00	SKU_A 5.00	SKU_A 5.00

ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24 REVISED: NOVEMBER 14 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 13

50.58 70.32	50.58 70.32	50.58 70.32	50.58 70.32	50.58 70.32
SKU_B 7.00 34.49 56.74	SKU_B 7.00 34.49 56.74	SKU_B 7.00 34.49 56.74	SKU_B 7.00 34.49 56.74	SKU_B 7.00 34.49 56.74
SKU_C 10.00 21.73 42.73	SKU_C 10.00 21.73 42.73	SKU_C 10.00 21.73 42.73	SKU_C 10.00 21.73 42.73	SKU_C 10.00 21.73 42.73

Table: Safety stock computation inputs (illustrative)

7. Conclusion

The proposed solution facilitates big data-driven analytics across AWS, Azure, and GCP, thereby improving the accuracy of supply chain optimization in the national food service wholesale sector, which includes supply, distribution, and economic supply service providers managed via the three major cloud service platforms. The federated capability addresses centralized data storage, potential data privacy disclosure when combining data in a third-party cloud, data governance across organizational boundaries, and latency issues in remote data access when combining supplier–distributor and distributor–store data to improve demand and inventory forecasting. Supported by the success of federated learning in machine learning model training, demand forecasting serves as analytics building block for data-hungry supply chain optimization problems and establishes a path to cross-cloud capabilities for other data-hungry algorithms needing supplier–distributor and distributor–store data combinations.

As new challenges arise in the future, such as climate change, persistent shortages of energy and raw materials, food security, diversity of origins, food safety, and consumer health demands, they must be examined in parallel with compliance with European regulations, reduction of greenhouse gas emissions, technological capabilities, national welfare, and degree of care taken by industry players. Technological adaptations such as big data, artificial intelligence, the Internet of Things, and cloud computing will foster breakthroughs and contribute to optimizing supply chains, logistics, transportation, and routing. Directions for future developments are detection and data renovation at the store level, integration and renovation by cloud services, study of full partnership and service relations among producers and transporters, effects of storage length, and means of assisting with supply and logistics costs.

7.1. Future Trends

Data-driven federated analytics has the potential to facilitate national supply chain optimization for food-service wholesale. The basic requirements for success are well-defined data governance and policies that govern source-data retention, sharing, and cross-cloud data location. The demand-forecasting, inventory-optimization, and transportation-and-logistics-routing modules can operate effectively in production with a few hundred SKU time's location combinations arranged for each week. The distribution division for each major area, namely, the West, East, Central, and Border regions, should contain independent suppliers that fulfill the interfacing requirement of low-latency development. The remaining supply-chain modules can be adapted progressively as the service demand grows or the data contribution burden can be shared with suppliers stationed in other major areas. With strong caution in applying data-derived results to decision making, federated analytics can enable strategic cloud-enabled national-level supply-chain decisions.

Looking beyond food service, federated analytics can be advantageous for national requirements in public health, epidemic control, pandemic containment, and climate change. These areas typically demand support from many cloud service platforms for properly managing shared data and workloads with low latency. However, such

requirements cannot be handled centrally due to numerous data-gathering points, resource-demanding high-precision prediction demands, and the data risk factor that tends to limit citizen participation in delivery. With proper policy support and technological empowerment, federated analytics can readily model the national service domain in public health and epidemic control by combining analytical models from different service providers.

8. References

[1] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Van Overveldt, T., Petrou, D., Ramage, D., & Roselander, J. (2019). Towards federated learning at scale: System design. In Proceedings of Machine Learning and Systems (MLSys).

[2] Dai, K., et al. (2024). A privacy-preserving multi-center federated learning framework for district heating forecast. Energy and Buildings, 328, 115164.

[3] Deng, S., Zhao, H., Huang, B., Zhang, C., Chen, F., Deng, Y., Yin, J., Dustdar, S., & Zomaya, A. Y. (2024). Cloud-native computing: A survey from the perspective of services. Proceedings of the IEEE, 112(1), 12–46.

[4] Dong, H., Zhang, C., Li, G., & Zhang, H. (2024). Cloud-native databases: A survey. IEEE Transactions on Knowledge and Data Engineering, 36(12), 7772–7798.

[5] Douaioui, K., Oucheikh, R., Benmoussa, O., & Mabrouki, C. (2024). Machine learning and deep learning models for demand forecasting in supply chain management: A critical review. Applied System Innovation, 7(5), 93.

[6] Fernández, J. D., et al. (2022). Privacy-preserving federated learning for residential short-term load forecasting. Applied Energy, 314, 118633.

[7] Guo, W., et al. (2024). A comprehensive survey of federated transfer learning. Frontiers of Computer Science, 18, 183101.

[8] Hübner, N., Caspers, J., Coroamă, V. C., & Finkbeiner, M. (2024). Machine-learning-based demand forecasting against food waste: Life cycle environmental impacts and benefits of a bakery case study. Journal of Industrial Ecology, 28(5), 1117–1131.

[9] Ibrahim Khalaf, O., et al. (2024). Federated learning with hybrid differential privacy for privacy–utility trade-offs. Security and Privacy, 7(2), e374.

[10] Jauhar, S. K., et al. (2024). Explainable artificial intelligence to improve the resilience of perishable product supply chains by leveraging customer characteristics. Annals of Operations Research.

[11] Jiang, S., et al. (2024). Fed-MPS: Federated learning with local differential privacy using model parameter selection for resource-constrained CPS. Journal of Systems Architecture, 150, 103108.

[12] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S.,

Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

[14] Khedr, A. M., & Rani, S. (2024). Enhancing supply chain management with deep learning and machine learning techniques: A review. Journal of Open Innovation: Technology, Market, and Complexity, 10(4), 100379.

[15] Kong, L., Zheng, G., & Brintrup, A. (2024). A federated machine learning approach for order-level risk prediction in supply chain financing. International Journal of Production Economics, 268, 109095.

[16] Li, J., et al. (2024). A comprehensive survey on client selection strategies in federated learning. Computer Networks, 242, 110028.

[17] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. In Proceedings of Machine Learning and Systems (MLSys).

[18] Liu, Y., Fan, L., Mao, Y., Chen, Z., & Li, S. (2025). A survey of optimization algorithms for differential privacy in federated learning. Computer Standards & Interfaces, 95, 103___.

[19] Maher, M., Oun, O. F., Mesmeh, M. S., & El Shawi, R. (2025). FedForecaster: An automated federated learning approach for time-series forecasting. In Proceedings of the 28th International Conference on Extending Database Technology (EDBT).

[20] Mathew, C., et al. (2024). Improved data privacy with differential privacy in federated learning. Journal of Web User Interface and Usability Analytics, 15(3), 1–___.

[21] Nikkhah, A., et al. (2024). Machine learning-based life cycle assessment for improving food supply chains toward sustainability. Sustainable Production and Consumption, 49, 1–___.

[22] Perifanis, V., Pavlidis, N., Koutsiamanis, R.-A., & Efraimidis, P. S. (2023). Federated learning for 5G base station traffic forecasting. Computer Networks, 227, 109950.

[23] Ranpara, R., et al. (2025). A semantic and ontology-based framework for enhancing interoperability and automation in IoT systems. Discover Internet of Things, 5, 122.

[24] Shadid, N., et al. (2025). A systematic review of data-driven approaches to food demand forecasting and waste reduction. Sustainable Futures, 7, 100___.

[25] Shan, F., et al. (2025). A survey of optimization algorithms for differential privacy in federated learning. Computer Standards & Interfaces, 95, 103___.

[26] Shi, C., et al. (2025). FedAWA: Adaptive optimization of aggregation weights in federated learning using client vectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[27] Stranieri, F., et al. (2024). Combining deep reinforcement learning and multi-stage

stochastic programming to address the supply chain inventory management problem. International Journal of Production Economics, 268, 109099.

[28] Tayyeh, H. K., et al. (2024). A differential privacy approach in federated learning. Computers, 13(11), 277.

[29] Thwal, C. M., et al. (2024). Transformers with attentive federated aggregation for time-series forecasting in non-IID settings. Expert Systems with Applications, 252, 123___.

[30] Wan, X., Yang, D., Wang, T., & Deveci, M. (2023). Closed-loop supply chain decision considering information reliability and security: Should the supply chain adopt federated learning decision support systems? Annals of Operations Research.

[31] Wei, J., et al. (2024). FedTWA: A federated learning model for supply chain demand prediction with temporal pattern attention. In Proceedings of the ACM International Conference on AI in Finance / Applied AI (ACM).

[32] Xia, J., et al. (2024). PT-ADP: A personalized privacy-preserving federated learning scheme based on transaction mechanism. Information Sciences, 669, 120519.

[33] Yang, F., et al. (2024). An explainable federated learning and blockchain-based secure credit modeling method. European Journal of Operational Research, 317(2), 449–467.

[34] Zheng, G., Kong, L., & Brintrup, A. (2023). Federated machine learning for privacy preserving, collective supply chain risk prediction. International Journal of Production Research, 61(23), 8115–8132.

[35] Zheng, G., Ivanov, D., & Brintrup, A. (2025). An adaptive federated learning system for information sharing in supply chains. International Journal of Production Research, 63(11), 3938–3960.

[36] Zogaan, W. A., et al. (2025). Leveraging deep learning for risk prediction and resilience in supply chains: A big data perspective. Journal of Big Data, 12, 143.


ISSN: 3067-283X
AMERICAN JOURNAL OF ANALYTICS AND ARTIFICIAL INTELLIGENCE
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 24
REVISED: NOVEMBER 14
ACCEPTED: NOVEMBER 28
PUBLISHED: DECEMBER 13