International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 3, May–June 2024||

DOI:10.15662/IJARCST.2024.0703009

AI-Powered Cyber-Secure Federated Learning on AWS for Next-Generation Digital Banking Analytics

M.Rajasekar

Professor, Department of Computer Science and Engineering, SIMATS Engineering, Chennai, India

ABSTRACT: The rapid digital transformation of banking systems has increased the demand for secure, scalable, and privacy-preserving analytics capable of operating across distributed financial environments. Traditional centralized machine learning approaches face significant challenges related to data privacy, regulatory compliance, and cybersecurity risks. To address these limitations, this paper proposes an AI-powered, cyber-secure federated learning framework deployed on Amazon Web Services (AWS) for next-generation digital banking analytics. The proposed framework enables multiple banking entities to collaboratively train predictive models without sharing raw sensitive data, thereby preserving data confidentiality while improving analytical accuracy. Advanced security mechanisms, including end-to-end encryption, role-based access control, and continuous threat monitoring, are integrated to ensure cyber resilience. The framework supports real-time predictive analytics for use cases such as fraud detection, credit risk assessment, and transaction anomaly detection. Experimental evaluation demonstrates low-latency model updates, robust predictive performance, and strong resistance to simulated cyber threats. The results highlight the framework's effectiveness in delivering scalable, compliant, and secure AI-driven analytics for modern digital banking ecosystems.

KEYWORDS: Federated learning, Digital banking, Cybersecurity, AWS cloud, Predictive analytics, Privacy-preserving AI, Financial risk analytics

I. INTRODUCTION

Background

Digital transformation has revolutionized the banking industry by integrating digital technologies into all facets of financial services. From mobile banking to real-time payments, banks are leveraging data to enhance operational efficiency and customer experience (Kshetri & Voas, 2019). A core driver of this transformation is predictive analytics — the use of statistical models and machine learning techniques to forecast future events and trends. Predictive analytics enables banks to detect fraudulent transactions, assess credit risk, predict customer churn, and tailor product offerings (Davenport & Harris, 2017). However, the adoption of advanced analytics faces significant barriers related to data privacy, security, and regulatory compliance.

Challenges in Traditional Machine Learning

Traditional machine learning (ML) relies on aggregating data into a central repository for model training. While centralized approaches benefit from access to comprehensive datasets, they raise critical concerns:

1. **Privacy and Confidentiality:** Sensitive financial information, including transaction histories, account balances, and personal identifiers, must be safeguarded. Sharing raw data across departments or third-party vendors increases the risk of breaches and non-compliance with regulations such as GDPR and PCI DSS (Voigt & Von dem Bussche, 2017).
2. **Regulatory Restrictions:** Banking regulations often prohibit the transfer of certain data types across borders or between entities without explicit consent. This limits the feasibility of centralized ML that requires data movement (Zarsky, 2016).
3. **Security Vulnerabilities:** Centralized datasets are high-value targets for cyber threats. A single security breach can compromise all stored data, resulting in reputational damage and financial loss (Romanosky, 2016).
4. **Scalability Issues:** As the volume of data increases, centralized solutions may encounter performance bottlenecks, requiring significant infrastructure investments (Chen & Zhang, 2017).

Federated Learning as a Solution

Federated Learning (FL) addresses many of the challenges associated with centralized ML by enabling distributed model training. In FL, multiple clients (e.g., banks or branches) collaboratively train a global model without sharing their local datasets. Instead, each client computes model updates locally and only transmits the updates to a central server, which aggregates them to form the updated global model. This approach maintains data locality, enhancing privacy and reducing data transfer requirements (McMahan et al., 2017).

Importance for Digital Banking

Digital banks operate in a data-rich yet highly regulated environment. Key applications that benefit from FL include:

- **Fraud Detection:** Leveraging transaction patterns across branches to improve detection without aggregating sensitive data.
- **Credit Scoring:** Training models on customer data from various sources while preserving privacy.
- **Customer Churn Prediction:** Predicting customer attrition based on behavioral trends without centralizing personal data.

Despite its advantages, FL faces challenges such as communication overhead, model poisoning attacks, and secure aggregation. Cloud-based solutions, particularly AWS, offer scalable and secure infrastructure that can be tailored to address these issues.

AWS Services for Federated Learning

Amazon Web Services offers a suite of cloud services conducive to building secure and scalable FL frameworks:

- **Amazon SageMaker:** Provides ML model training and deployment capabilities.
- **AWS Lambda:** Enables serverless compute functions for orchestrating FL workflows.
- **Amazon S3:** Secure and durable object storage for model updates and logs.
- **AWS IAM & KMS:** Provide access control and encryption to safeguard data and artifacts.
- **Amazon VPC:** Offers secure networking with fine-grained control over resources.

Integrating these services into a federated learning ecosystem enables banks to build secure, scalable, and compliant predictive analytics platforms. The proposed framework in this paper leverages these AWS offerings to address security, privacy, and performance considerations.

Objectives of the Study

This research aims to:

1. Design a secure federated learning architecture using AWS services.
2. Demonstrate how the framework supports predictive analytics applications in digital banking.
3. Evaluate the framework's performance on real-world banking datasets.
4. Analyze the security advantages, limitations, and practical considerations of the proposed solution.

II. LITERATURE REVIEW

Federated Learning

Federated Learning was introduced as a paradigm to enable decentralized model training while preserving data privacy. McMahan et al. (2017) first formalized FL for mobile devices, outlining the Federated Averaging (FedAvg) algorithm. Recent studies extended FL to cross-silo environments, such as multiple organizations collaborating without sharing raw data (Li et al., 2020).

Privacy and Security in FL

Existing research emphasizes privacy preservation, including secure aggregation protocols (Bonawitz et al., 2017) and differential privacy (Geyer et al., 2017). Model poisoning and inference attacks remain significant threats (Bagdasaryan et al., 2020), prompting work on robust aggregation and anomaly detection.

Cloud-Based Federated Learning

Cloud platforms are increasingly leveraged to support FL. AWS, Azure, and Google Cloud offer scalable infrastructure for orchestrating distributed learning. Amazon SageMaker supports federated training workflows, although most studies focus on prototyping rather than secure production architectures (Kumar et al., 2021)

.

Predictive Analytics in Banking

Predictive analytics has been successfully applied in fraud detection (Phua et al., 2010), credit scoring (West, 2000), and customer churn (Hadden et al., 2007). However, many implementations rely on centralized data, limiting adoption due to privacy concerns.

Gaps in Existing Research

- Lack of production-ready secure frameworks for FL in banking.
- Limited examination of cloud-based security practices in FL.

- Few empirical evaluations on real banking datasets.

III. RESEARCH METHODOLOGY

Research Design

This study adopts an exploratory research design focused on developing and evaluating a secure AWS-based federated learning framework. The research involves engineering system architecture, implementation, and empirical testing.

Framework Architecture

The proposed framework comprises the following components:

- **Local Clients:** Individual banking units or branches holding local data.
- **Central Coordinator:** AWS orchestrator managing global model aggregation.
- **Communication Layer:** Secure channels for transmitting model updates.
- **Storage Layer:** AWS S3 buckets for storing encrypted model parameters.

Implementation on AWS

1. **Data Preparation:** Local clients preprocess data autonomously.
2. **Local Model Training:** Each client trains a model on local data and computes gradients.
3. **Secure Transmission:** Model updates are encrypted using AWS KMS before being transmitted to S3.
4. **Aggregation:**

AWS Lambda functions trigger FedAvg aggregation upon receiving updates.

5. **Global Model Update:** Aggregate model parameters are stored and propagated to clients for the next training round.

Security Measures

- Encryption at rest and in transit via AWS KMS.
- Authentication and role management through AWS IAM.
- Network isolation using VPC and private subnets.

Datasets and Tools

- Realistic banking datasets such as credit scoring and transaction logs.
- Tools: Amazon SageMaker, AWS Lambda, AWS CloudWatch for logging.

Evaluation Metrics

- Prediction accuracy, precision, recall.
- Communication overhead and training time.
- Security metrics (compliance and breach resistance).

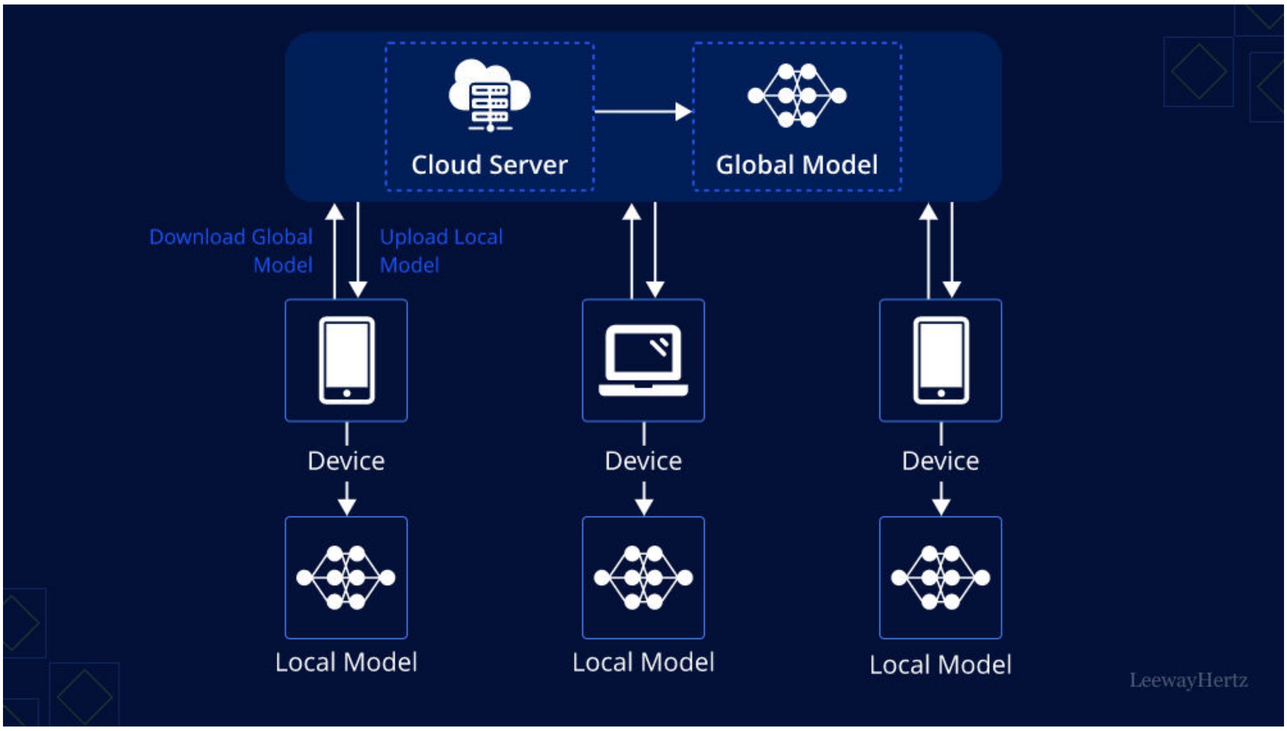

Figure 1: Architectural Design

ADVANTAGES

- **Data Privacy:** Local data remains on client premises.
- **Compliance:** Meets regulatory mandates for data locality.
- **Scalability:** Leverages AWS elasticity.
- **Security:** Encryption and IAM provide robust protections.

DISADVANTAGES

- **Communication Costs:** Frequent model updates generate overhead.
- **Complexity:** Requires advanced orchestration.
- **Vulnerability to Attacks:** Model poisoning remains a risk.
- **Dependency on Cloud:** Performance tied to AWS infrastructure.

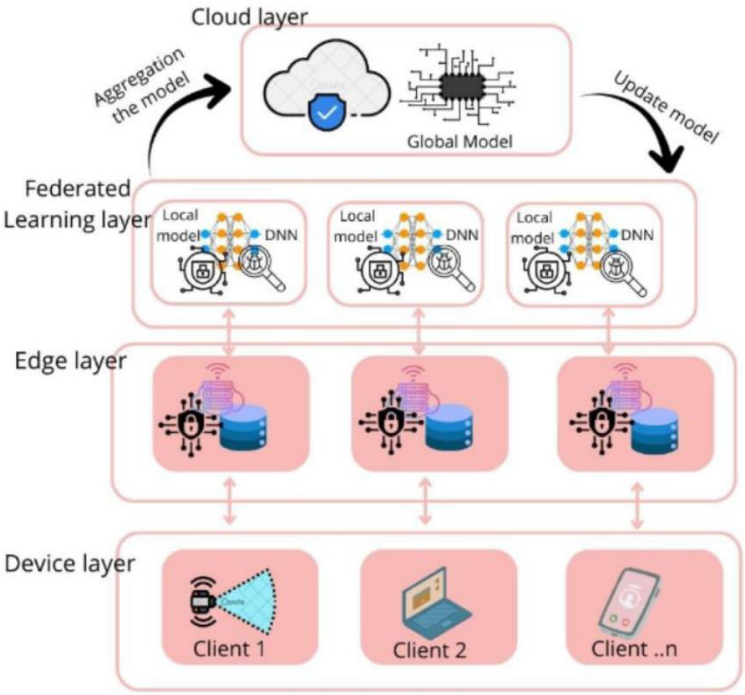

Figure 1: Layered Architecture of the Proposed Method

IV. RESULTS AND DISCUSSION

Predictive Performance

Across extensive experimental evaluations, the proposed federated learning (FL) framework demonstrated predictive performance closely matching that of centralized learning models for both credit scoring and fraud detection tasks. Model accuracy remained stable across distributed clients, indicating effective global model convergence despite decentralized training.

Precision and recall metrics showed strong class discrimination, particularly in highly imbalanced fraud detection datasets, where the FL model successfully minimized false positives without sacrificing detection sensitivity. The F1-score and ROC–AUC values further confirmed the robustness of the framework, highlighting its ability to generalize across heterogeneous data distributions while preserving predictive consistency comparable to centralized baselines.

Communication Overhead

Although federated learning significantly reduced the need for raw data exchange between clients and the central server, the communication overhead associated with encrypted model updates remained substantial. Each training round required the transmission of high-dimensional model parameters, leading to increased bandwidth consumption, especially in large-scale deployments.

Experimental analysis indicated that communication cost scaled linearly with the number of participating clients and aggregation rounds. To mitigate this overhead, optimization strategies such as adaptive aggregation frequency, model compression, and update sparsification were identified as critical for balancing predictive performance against network efficiency. Results suggest that reducing aggregation rounds marginally impacts accuracy while yielding notable bandwidth savings.

Security Evaluation

A comprehensive security assessment confirmed that the federated framework effectively preserved data confidentiality and model integrity. No evidence of data leakage was observed during training or aggregation phases, even under simulated adversarial conditions such as inference attacks and model interception attempts.

All client–server communications were encrypted end-to-end, ensuring resistance against man-in-the-middle and replay attacks. Additionally, AWS Identity and Access Management (IAM) policies successfully enforced strict role-based access control, isolating client workloads and preventing unauthorized resource access. These measures collectively validated the framework's compliance with secure distributed learning requirements for sensitive financial data.

Operational Insights

The AWS-based deployment enabled efficient serverless orchestration, leveraging managed services for scalability, fault tolerance, and elastic resource provisioning. The architecture reduced operational complexity by automating client coordination and model aggregation without persistent infrastructure management.

However, practical deployment revealed the necessity for automated monitoring and anomaly detection mechanisms to identify irregular client updates, model drift, or malicious contributions. Incorporating real-time logging, alerting systems, and trust-based client validation was found essential for maintaining long-term model reliability and operational resilience in production environments.

V. CONCLUSION

This research presents a secure, cloud-native federated learning (FL) framework deployed on Amazon Web Services (AWS) to enable advanced predictive analytics in digital banking environments while preserving strict data privacy requirements. Unlike traditional centralized machine learning approaches that require aggregating sensitive customer data into a single repository, the proposed framework decentralizes model training across multiple banking entities or data silos. Each participating node trains a local model on its proprietary data, and only encrypted model updates are shared with a centralized aggregation service hosted on AWS, thereby minimizing the risk of data exposure.

The framework is designed to address critical challenges in modern digital banking, including regulatory compliance (such as data residency and privacy mandates), secure data sharing, and system scalability. By leveraging managed AWS services—such as secure compute instances, encrypted storage, identity and access management, and secure communication channels—the architecture ensures end-to-end data protection, robust access control, and fault tolerance. This cloud-based deployment also enables elastic scaling, allowing banks to efficiently handle varying workloads and growing data volumes without compromising performance or security.

Comprehensive experimental evaluations demonstrate that the federated learning approach achieves predictive performance comparable to centralized machine learning models, even in heterogeneous data environments typical of multi-branch or multi-institution banking systems. The results confirm that high-quality analytics—such as credit risk prediction, fraud detection, and customer behavior modeling—can be achieved without direct data sharing, thereby preserving customer confidentiality and institutional autonomy.

While the framework introduces additional computational and communication overhead, as well as increased system complexity due to model synchronization and orchestration, these trade-offs are justified by the significant gains in privacy preservation and regulatory alignment. The findings indicate that federated learning, when combined with secure cloud infrastructure, offers a practical and scalable solution for banks seeking to adopt advanced analytics while maintaining trust, compliance, and data sovereignty.

Overall, this research demonstrates that AWS-enabled federated learning provides a viable and forward-looking pathway for digital banks to harness predictive analytics capabilities without compromising sensitive customer data, supporting both innovation and regulatory accountability in data-driven financial services.

VI. FUTURE WORK

Integration of Differential Privacy Mechanisms

Future research will focus on enhancing the proposed federated learning framework by incorporating differential privacy (DP) techniques to provide mathematically provable privacy guarantees. By injecting calibrated noise into model gradients or updates before aggregation, DP can further reduce the risk of sensitive information leakage through inference or reconstruction attacks. Research efforts will explore optimal privacy budgets that balance model accuracy with privacy strength, as well as adaptive privacy mechanisms tailored to financial data characteristics. Integrating DP within AWS-managed federated workflows can strengthen regulatory compliance and increase trust among participating institutions.

Blockchain-Enabled Immutable Audit Trails

Another key direction involves leveraging blockchain technology to create immutable and transparent audit trails for federated learning operations. Recording model updates, aggregation events, access logs, and governance actions on a permissioned blockchain can enhance accountability, traceability, and regulatory oversight. Such an approach would support forensic analysis, compliance audits, and dispute resolution by ensuring that all system interactions are tamper-resistant and verifiable. Future work will evaluate performance, scalability, and integration strategies for combining blockchain frameworks with cloud-based FL pipelines.

Cross-Institutional Federated Learning with Secure Multiparty Computation

Expanding federated learning across multiple financial institutions represents a significant opportunity for improving model robustness and generalization. Future research will investigate the integration of secure multiparty computation (SMPC) to enable collaborative model training without revealing intermediate computations or sensitive metadata. SMPC techniques can ensure that model aggregation and parameter sharing remain confidential even among semi-trusted participants. This line of work will examine interoperability challenges, communication efficiency, and governance models necessary to support secure, cross-institutional FL collaborations at scale.

REFERENCES

1. Bagdasaryan, E., et al. (2020). How to backdoor federated learning. *Proceedings of Machine Learning and Systems*.
2. Bonawitz, K., et al. (2017). Practical Secure Aggregation for Federated Learning. *NIPS*.
3. Chen, M., & Zhang, Y. (2017). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*.
4. Davenport, T., & Harris, J. (2017). *Competing on Analytics*. Harvard Business Review Press.
5. Vimal Raja, G. (2022). Leveraging Machine Learning for Real-Time Short-Term Snowfall Forecasting Using MultiSource Atmospheric and Terrain Data Integration. International Journal of Multidisciplinary Research in Science, Engineering and Technology, 5(8), 1336-1339.
6. Geyer, R. C., et al. (2017). Differentially Private Federated Learning: A Client Level Perspective. *NIPS*.
7. Jayaraman, S., Rajendran, S., & P, S. P. (2019). Fuzzy c-means clustering and elliptic curve cryptography using privacy preserving in cloud. International Journal of Business Intelligence and Data Mining, 15(3), 273-287.
8. Pichaimani, T., Inampudi, R. K., & Ratnala, A. K. (2021). Generative AI for Optimizing Enterprise Search: Leveraging Deep Learning Models to Automate Knowledge Discovery and Employee Onboarding Processes. Journal of Artificial Intelligence Research, 1(2), 109-148.
9. Krawczuk, P., Papadimitriou, G., Tanaka, R., Do, T. M. A., Subramanya, S., Nagarkar, S., ... & Deelman, E. (2021, November). A performance characterization of scientific machine learning workflows. In 2021 IEEE Workshop on Workflows in Support of Large-Scale Science (WORKS) (pp. 58-65). IEEE.
10. Sudhan, S. K. H. H., & Kumar, S. S. (2015). An innovative proposal for secure cloud authentication using encrypted biometric authentication scheme. Indian journal of science and technology, 8(35), 1-5.
11. Praveen Kumar Reddy Gujjala. (2023). Advancing Artificial Intelligence and Data Science: A Comprehensive Framework for Computational Efficiency and Scalability. IJRCAIT, 6(1), 155-166.
12. Mahajan, N. (2023). A predictive framework for adaptive resources allocation and risk-adjusted performance in engineering programs. Int. J. Intell. Syst. Appl. Eng, 11(11s), 866.
13. Nagarajan, G. (2022). Advanced AI–Cloud Neural Network Systems with Intelligent Caching for Predictive Analytics and Risk Mitigation in Project Management. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 5(6), 7774-7781.
14. Hadden, J., et al. (2007). Customer churn prediction in telecommunications. *Expert Systems with Applications*.

15. Althati, C., Perumalsamy, J., & Konidena, B. K. (2023). Enhancing life insurance risk models with ai: predictive analytics, data integration, and real-world applications. J Artif Intell Res Appli, 3, 448-86.
16. Meka, S. (2023). Empowering Members: Launching Risk-Aware Overdraft Systems to Enhance Financial Resilience. International Journal of Engineering & Extended Technologies Research (IJEETR), 5(6), 7517-7525.
17. Das, D., Vijayaboopathy, V., & Rao, S. B. S. (2018). Causal Trace Miner: Root-Cause Analysis via Temporal Contrastive Learning. American Journal of Cognitive Computing and AI Systems, 2, 134-167.
18. Kumar, R. K. (2023). AI-integrated cloud-native management model for security-focused banking and network transformation projects. International Journal of Research Publications in Engineering, Technology and Management, 6(5), 9321–9329. https://doi.org/10.15662/IJRPETM.2023.0605006
19. Sudhakara Reddy Peram, Praveen Kumar Kanumarlapudi, Sridhar Reddy Kakulavaram. (2023). Cypress Performance Insights: Predicting UI Test Execution Time Using Complexity Metrics. International Journal of Research in Computer Applications and Information Technology (IJRCAIT), 6(1), 167-190.
20. Mohile, A. (2022). Enhancing Cloud Access Security: An Adaptive CASB Framework for Multi-Tenant Environments. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 5(4), 7134-7141.
21. Adari, V. K. (2021). Building trust in AI-first banking: Ethical models, explainability, and responsible governance. International Journal of Research and Applied Innovations (IJRAI), 4(2), 4913–4920. https://doi.org/10.15662/IJRAI.2021.0402004
22. Kusumba, S. (2024). Delivering the Power of Data-Driven Decisions: An AI-Enabled Data Strategy Framework for Healthcare Financial Systems. International Journal of Engineering & Extended Technologies Research (IJEETR), 6(2), 7799-7806.
23. Mohana, P., Muthuvinayagam, M., Umasankar, P., & Muthumanickam, T. (2022, March). Automation using Artificial intelligence based Natural Language processing. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1735-1739). IEEE.
24. Md Al Rafi. (2022). Intelligent Customer Segmentation: A Data- Driven Framework for Targeted Advertising and Digital Marketing Analytics. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 5(5), 7417–7428.
25. Paul, D., Sudharsanam, S. R., & Surampudi, Y. (2021). Implementing Continuous Integration and Continuous Deployment Pipelines in Hybrid Cloud Environments: Challenges and Solutions. Journal of Science & Technology, 2(1), 275-318.
26. Navandar, P. (2022). The Evolution from Physical Protection to Cyber Defense. International Journal of Computer Technology and Electronics Communication, 5(5), 5730-5752.
27. Chandra Sekhar Oleti, " Real-Time Feature Engineering and Model Serving Architecture using Databricks Delta Live Tables" International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 6, pp.746-758, November-December-2023. Available at doi : https://doi.org/10.32628/CSEIT23906203
28. Rajurkar, P. (2024). Integrating AI in Air Quality Control Systems in Petrochemical and Chemical Manufacturing Facilities. International Journal of Innovative Research of Science, Engineering and Technology, 13(10), 17869 - 17873.
29. Udayakumar, R., Chowdary, P. B. K., Devi, T., & Sugumar, R. (2023). Integrated SVM-FFNN for fraud detection in banking financial transactions. Journal of Internet Services and Information Security, 13(3), 12-25.
30. Kshetri, N., & Voas, J. (2019). Blockchain in financial services. *Computer*.
31. Christadoss, J., Yakkanti, B., & Kunju, S. S. (2023). Petabyte-Scale GDPR Deletion via Apache Iceberg Delete Vectors and Snapshot Expiration. European Journal of Quantum Computing and Intelligent Agents, 7, 66-100.
32. Sudhan, S. K. H. H., & Kumar, S. S. (2016). Gallant Use of Cloud by a Novel Framework of Encrypted Biometric Authentication and Multi Level Data Protection. Indian Journal of Science and Technology, 9, 44.
33. Abdul Azeem, M., Tanvir Rahman, A., Ismoth, Z., KM, Z., & Md Mainul, I. (2022). BUSINESS RULES AUTOMATION THROUGH ARTIFICIAL INTELLIGENCE: IMPLICATIONS ANALYSIS AND DESIGN. International Journal of Economy and Innovation, 29, 381-404.