Contribution ID: **57** Type: **Oral**

Enabling Trustworthy Big Data Analytics through Federated Learning

The exponential growth of data in sectors such as healthcare, finance, and smart infrastructure has underscored the transformative potential of Big Data analytics. However, centralizing sensitive information raises concerns related to privacy, security, and regulatory compliance, making traditional machine learning approaches increasingly unsuitable. To overcome these limitations, this research advances Privacy-Preserving Big Data Analytics using Federated Learning (FL), a decentralized paradigm where models are collaboratively trained across distributed nodes without transferring raw data.
The study proposes scalable FL architectures specifically designed for heterogeneous and large-scale datasets. To safeguard against adversarial threats and data leakage, the framework integrates differential privacy and secure aggregation techniques. Furthermore, it addresses practical challenges such as non-IID data distributions, communication overhead, and slow convergence, which often limit the effectiveness of federated deployments in real-world environments. The proposed framework has broad applicability in healthcare diagnostics, IoT ecosystems, and smart city solutions, where ensuring confidentiality while maintaining predictive accuracy is critical. The expected outcomes include the development of a robust, adaptable, and ethically responsible Big Data analytics framework that balances scalability, security, and performance. By bridging the gap between Big Data and privacy-preserving distributed learning, this research contributes to the advancement of trustworthy, next-generation AI systems.

Keywords: Big Data Analytics, Federated Learning, Privacy-Preserving Machine Learning, Differential Privacy, Secure Aggregation, Trustworthy AI

Author: V, Arun (Jain University)

Presenter: V, Arun (Jain University)

Track Classification: Mathematical & Data Sciences