

2nd International Conference on Computing and Machine Learning (CML 2025)

Mr. Sunil Kumar Mudusu

Talk Title:

The Importance of Data Quality in Machine Learning: Challenges and Solutions

Abstract / Profile:

Abstract: Data quality is a fundamental determinant of machine learning success, influencing model accuracy, reliability, and overall business impact. Poor data quality can result in biased models, incorrect predictions, and suboptimal decision-making, leading to financial and operational setbacks. Organizations face numerous challenges in ensuring high data quality, including data silos, bias, scalability limitations, and insufficient governance. This paper examines these challenges and explores the data quality lifecycle, covering acquisition, cleaning, validation, and monitoring. Advanced techniques such as machine learning-powered validation, automated profiling, and synthetic data generation are also discussed. Additionally, fostering a data quality-driven culture through leadership commitment, employee training, and clear KPIs is essential for sustainable data integrity. Looking ahead, innovations like blockchain-enabled data provenance, AI-driven governance, and federated learning will further enhance data quality management. By prioritizing data quality, organizations can unlock the full potential of machine learning, leading to more accurate insights and better business outcomes.