# **Distributed Machine Learning (DML): Partitioning Techniques in AI Models for Wireless Network Optimization**

|  |  |  |
| --- | --- | --- |
| **Name** | **BITS ID** | **Contribution** |
| SUBHRANSU MISHRA | 2023AC05489 | 100% |
| DULAL DAS | 2023AC05041 | 100% |
| LAKSHMISRINIVAS PERAKAM | 2023AC05540 | 100% |
| ARCHAN GHOSH | 2023AC05042 | 100% |

# Definitions in AI Model Design

# Vertical Partitioning

## **Definition**: Vertical partitioning involves splitting the features (columns) of a dataset across different sub-models or nodes.

## **Context in AI**: Each node or sub-model receives a subset of features for the same set of samples. This is common in **cross-silo federated learning**, where institutions (e.g., hospitals, telecom providers) hold different types of data about the same users.

## **Example**: In a wireless network, one node might process signal strength and frequency features, while another handles user mobility and device type.

# Horizontal Partitioning

## **Definition**: Horizontal partitioning splits the **samples (rows)** of a dataset across different nodes or sub-models.

## **Context in AI**: Each node trains on a different subset of users or devices, but with the same feature space. This is typical in **cross-device federated learning**.

## **Example**: In a wireless network, one node might train on data from urban users, another on rural users, each with full feature sets like signal strength, latency, and throughput.

# Comparison Across Key Dimensions

|  |  |  |
| --- | --- | --- |
| **Dimension** | **Vertical Partitioning** | **Horizontal Partitioning** |
| **Computational Efficiency** | Requires coordination across feature dimensions. May involve more complex joins | More efficient per node since each has full feature set. Easier parallel training |
| **Scalability** | Limited by feature diversity and inter-feature dependencies | Highly scalable across devices or regions. Ideal for large user bases |
| **Deployment in Wireless Networks** | Useful when different network components collect different features. Require secure feature sharing | Easier to deploy across distributed edge devices. Compatible with mobile and IoT scenarios |

## **Examples Using Wireless Network Datasets**

## **Dataset References**

## **SNDLib & Abilene Network**: Provide traffic matrices and topologies useful for partitioning experiments.

## **WACA (Wi-Fi All-Channel Analyzer)**: Captures RSSI across all Wi-Fi channels, ideal for vertical partitioning where each channel is a feature.

## **Google Stadia Traffic Dataset**: Useful for horizontal partitioning based on user behavior or resolution preferences.

## **Real-world RF Dataset (IEEE 802.11ax, LTE, 5G-NR)**: Contains signal traces across technologies, enabling both vertical (modality-based) and horizontal (user-based) partitioning.

## **Use Case Examples**

## **Vertical**: A telecom provider splits features between signal quality metrics and user mobility patterns, training sub-models separately and merging predictions.

## **Horizontal**: A federated learning setup where base stations train models on local user data (e.g., throughput, latency), then aggregate updates centrally.

## **Key Papers to Explore**

## **1.** [**Cross-Silo Federated Learning for Multi-Tier Networks**](https://arxiv.org/html/2108.08930v4)

## **Authors**: Anirban Das, Timothy Castiglia, Shiqiang Wang, Stacy Patterson

## **Focus**: Proposes a *Tiered Decentralized Coordinate Descent (TDCD)* algorithm for federated learning across silos with both vertical and horizontal partitioning.

## **Relevance**: Directly addresses wireless network scenarios with distributed data across clients and hubs.

## **2.** [**Vertical and Horizontal Partitioning in Data Stream Regression Ensembles**](https://jpbarddal.github.io/assets/pdf/vertical_horizontal_regression.pdf)

## **Author**: Jean Paul Barddal

## **Focus**: Investigates how combining vertical and horizontal partitioning improves regression accuracy in data stream mining.

## **Relevance**: Useful for understanding ensemble learning in dynamic wireless environments.

## **3.** [**Overview of Horizontal and Vertical Partitioning**](https://www.researchgate.net/profile/Mansi-Bosamia/publication/325570099_Overview_of_Horizontal_Partitioning_and_Vertical_Partitioning/links/5b167bf7a6fdcc31bbf5a2e6/Overview-of-Horizontal-Partitioning-and-Vertical-Partitioning.pdf)

## **Authors**: Meeta J. Pajwani, Mansi P. Bosamia

## **Focus**: Offers a foundational overview of partitioning strategies with implications for query optimization and distributed systems.

## **Relevance**: Good for conceptual grounding before diving into AI-specific applications.

## **4.** [**GeeksforGeeks Comparison Article**](https://www.geeksforgeeks.org/system-design/vertical-partitioning-vs-horizontal-partitioning/)

## **Title**: Vertical Partitioning vs Horizontal Partitioning

## **Focus**: A practical comparison of the two methods in system design.

## **Relevance**: While not a formal paper, it’s a helpful primer for understanding trade-offs.