Swarm Intelligence for Big Data Clustering and Optimization

The rapid expansion of big data has introduced new challenges in data clustering, feature selection, and optimization, where traditional algorithms often struggle with scalability and efficiency. Swarm intelligence (SI), a nature-inspired computational paradigm based on decentralized, self-organizing systems, has demonstrated significant potential in addressing these challenges. In this project, we focus on the application of swarm-based metaheuristic algorithms, such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Artificial Bee Colony (ABC), for clustering and feature selection in large-scale datasets. Recent studies (e.g., Yang et al., 2023; Zhang & Li, 2024) highlight the effectiveness of SI algorithms in handling high-dimensional data, optimizing model performance, and reducing computational costs.

To implement and evaluate these approaches, we will use Python along with machine learning and distributed computing libraries such as Scikit-learn, Dask, and PySpark. Experiments will be conducted on benchmark big data datasets, including those from the UCI Machine Learning Repository and real-world sources such as financial and biomedical data. Performance will be measured in terms of clustering accuracy, feature reduction efficiency, and computational speed, comparing SI-based methods against traditional clustering techniques like K-Means, DBSCAN, and spectral clustering. The expected outcome is a demonstration of how swarm intelligence can enhance data-driven decision-making by improving processing efficiency and adaptability in big data environments.