

A Survey on Scalable Feature Engineering Techniques for Cloud-Native Machine Learning Workflows

Raviteja Narra

Independent Researcher

ravitejanarra563@gmail.com

Abstract: *The presented survey provides a briefing of scalable feature engineering techniques of cloud-native machine learning (ML) processes based on the findings of the recent literature. It explores distributed feature extraction, large-scale feature transformations, feature selection, dimensionality reduction and feature representation using deep learning. The automated practices include stores, real-time streaming, feature versioning, feature lineage and feature governance are explored with respect to their perceived effects on the activity of reproducibility, consistency and scalability. The literature presents high-dimensional skewed data, computational constraints, orchestration complexity, and low-latency as some of the key Challenges in the field. The survey gives an overview of commerce offs between throughput and latency and resource efficiency and discusses integration approaches to scattered processing systems using cloud-native achievement systems. The observations can be utilized during the designing of adaptive feature engineering pipeline, resistant, and efficient within the modern cloud environment.*

Keywords: Scalable Feature Engineering, Cloud-Native Machine Learning, Automated Feature Engineering, Real-Time Feature Engineering, Resource Management, Dimensionality Reduction, Deep Learning Feature Representation