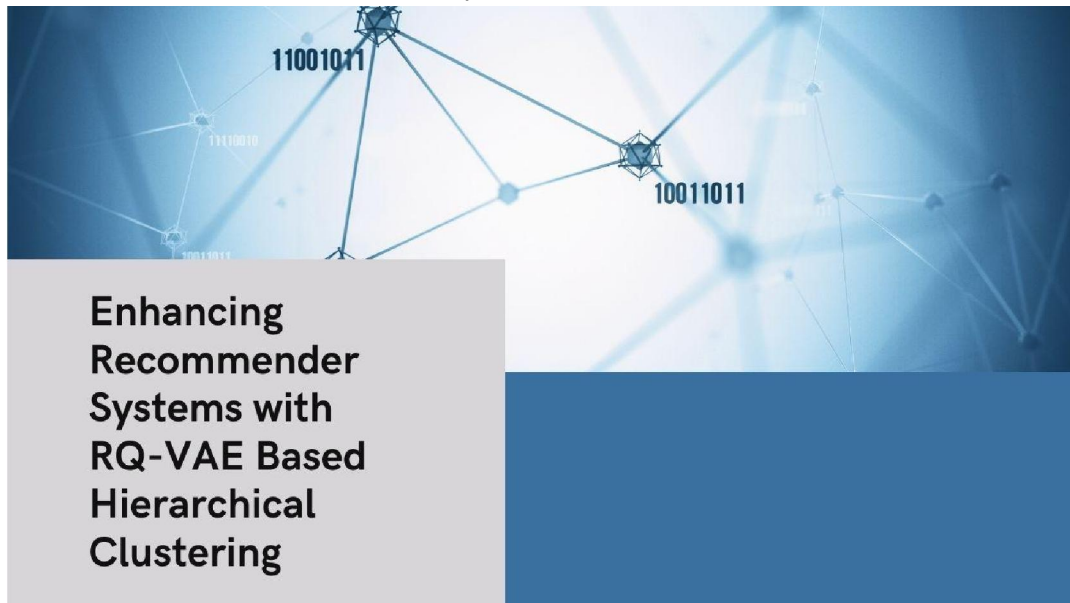# Enhancing Recommender Systems with RQ-VAE-Based Hierarchical Clustering

Sri Santhosh Hari
University of San Francisco, USA

Abstract: *In large-scale recommender systems, high-cardinality features pose significant challenges, including sparse representations, cold-start problems, limited generalization capabilities, and computational inefficiency. This article explores how the Residual-Quantized Variational AutoEncoder (RQ-VAE) addresses these challenges through hierarchical semantic IDs. The technique quantizes embeddings via a multi-level process that preserves semantic relationships while reducing dimensionality. By implementing residual connections between quantization levels, RQ-VAE creates a structured discrete representation where proximity in codeword sequences corresponds to item similarity. This article maintains the advantages of discrete representations while preserving semantic richness, enabling more effective cold-start handling, improved retrieval efficiency, and enhanced prediction performance. The hierarchical nature of semantic IDs facilitates generalization across similar items even with limited observation data, addressing fundamental limitations of traditional feature hashing and embedding-based methods while significantly improving recommendation quality for new and sparse items*