CSI3019 – ADVANCED DATA COMPRESSION TECHNIQUES

Team Members:
R S Shreenidhi-22MID0306
Keerthana G – 22MID0144
R Sivani – 22MID0005
Alyssa Mariam - 22MID0076

Title: **SEMANTIC-AWARE ARCHIVAL OF LOW-IMPACT SOCIAL MEDIA POSTS USING TIERED COMPRESSION AND COLD STORAGE**

Abstract:
With the explosive growth of social media platforms, vast amounts of user-generated content are created daily, much of which becomes obsolete within a short time. Storing all posts indefinitely places a heavy burden on infrastructure and storage costs, especially when a significant portion of the data is rarely accessed or used. This project proposes an intelligent, semantic-aware archival system that identifies low-impact social media posts based on engagement metrics and contextual relevance, then compresses and migrates them to cold storage using adaptive compression techniques. Posts are categorized into hot, warm, or cold tiers depending on their calculated impact score, where each tier applies a distinct compression strategy optimized for storage efficiency and access latency. The system retains lightweight metadata for quick retrieval, enabling selective decompression only when necessary. This approach ensures reduced storage overhead, improved scalability, and efficient data lifecycle management, making it a viable solution for big data platforms managing long-term social media archives.

Each post is analyzed and classified into three tiers:

Tier	Criteria (Example)	Access Frequency	Importance
Hot	Viral or recent posts with high engagement	High	Critical
Warm	Moderately engaging or trending posts	Medium	Moderate
Cold	Old, low-engagement posts	Low	Low