CSI3019-Advanced Compression Techniques

Review – 1
Emoji-Based Text Compression

Team Members:

DhivyaShree V–22MID0295
Jai Balaje S – 22MIC0123
Guruprasath S P – 22MID0067
Gnaneshwar A –22MID0224
Mohan Babu A-22MID0329

Abstract:

This project presents a novel, human-readable text compression technique for conversational chat data using emoji-based phrase substitution. The system employs a hybrid dictionary approach—combining a manually designed seed dictionary with a dynamically generated phrase dictionary—to replace frequently occurring chat expressions with semantically meaningful emojis. This method reduces storage and transmission overhead while ensuring complete reversibility and maintaining message interpretability. Experimental evaluation on real-world chat datasets demonstrates that the proposed approach achieves higher compression ratios than traditional character-level algorithms, making it suitable for live messaging platforms and large-scale chat archives.