LLM Performance Comparison on Code Generation Similarity Metrics