

Data Enrichment Using LLM Text Analytics

Pragyansmita Nayak, Ph.D.

Chief Data Scientist

Hitachi Vantara Federal

Text Analytics is a branch of machine learning (ML) combined with natural language processing (NLP) which extracts key characteristics from unstructured text documents. The additional processing cycles can improve trust for data assets by expanding their accessibility. Improved search of the content through correlation statistics, sentiment analysis, entity recognition and summarization can be easily automated.

Using Large Language Models (LLMs) such as OpenAI's ChatGPT, this talk will explore the potential data engineering operationalization and its impact on the associated metadata management. Keywords: LLM, ChatGPT, Text Analytics, Data Engineering, Unstructured Data, Sentiment Analysis, Summarization, Metadata Management, Feature Engineering, Hitachi Lumada DataOps.