Confidently design and apply algorithms for manipulating data in programming solutions for a variety of problems.

Our project Just Pick demonstrates use of algorithms to solve the problem of movie recommendation by integrating a local LLaMA language model with a vector-based semantic search.

We began by selecting dataset from Kaggle that included metadata such as keywords, descriptions, and cast information. Our focus was on efficiently organizing and preparing the data, this involved automating data ingestion, normalizing relational structures, and optimizing loading processes. We designed a normalized relational schema in MySQL to organize the movie data and support efficient querying. CSV-loading method was replaced with index file and a JSON metadata dump, enabling fast loading without rebuilding the recommendation index each time. We automated the process of loading raw CSV movie data into MySQL database.

We used chunk\_text method to split long movie descriptions into smaller segments, allowing the embedding model to process and generate accurate embeddings for each movie. We used Sentence Transformer to capture the semantic meaning of each movie. These embeddings were then indexed using FAISS, allowing fast searches based on user inputs. Relevant movies are retrieved and passed to a local LLaMA using prompt, which returns list of movie recommendations.