



Harnessing Vector Databases for AI and Data Search

Author: Michael Nguyen

Goal

Enable fast and intelligent semantic search over unstructured data using vector databases and AI embeddings.

Motivation

Traditional databases are built for structured data and exact matches—struggle to understand the *meaning* behind unstructured content like text, images, or user behavior.

- This limits their ability to reveal patterns, relationships, and context across diverse data *types*.
- As a result, users often miss key insights or waste time manually filtering through irrelevant results.
- Current tools that support semantic search are often costly, complex, or lack flexibility.

This project leverages **vector databases** to close that gap—delivering fast, intelligent, and meaning-aware search capabilities tailored for modern AI applications.

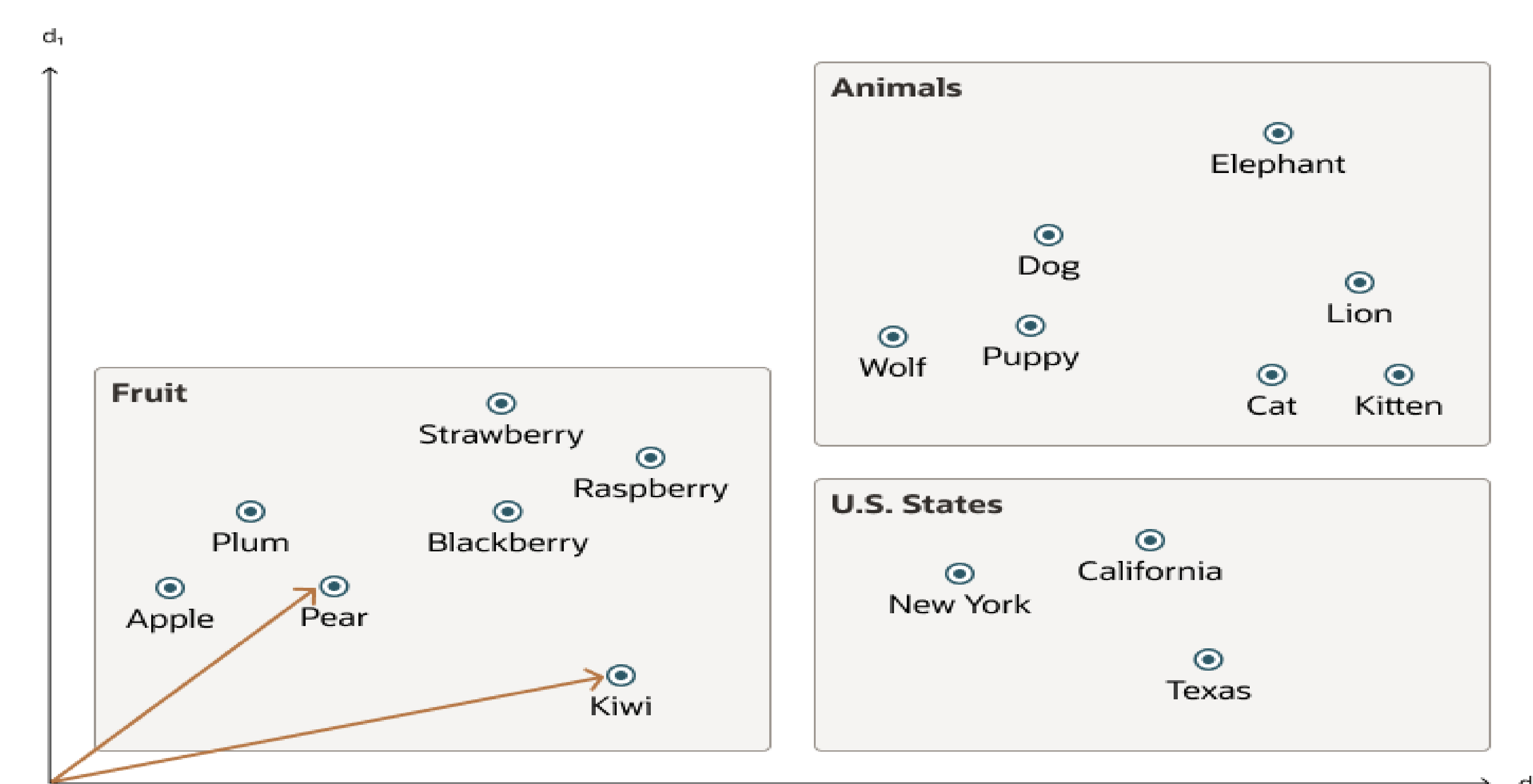
Project Summary



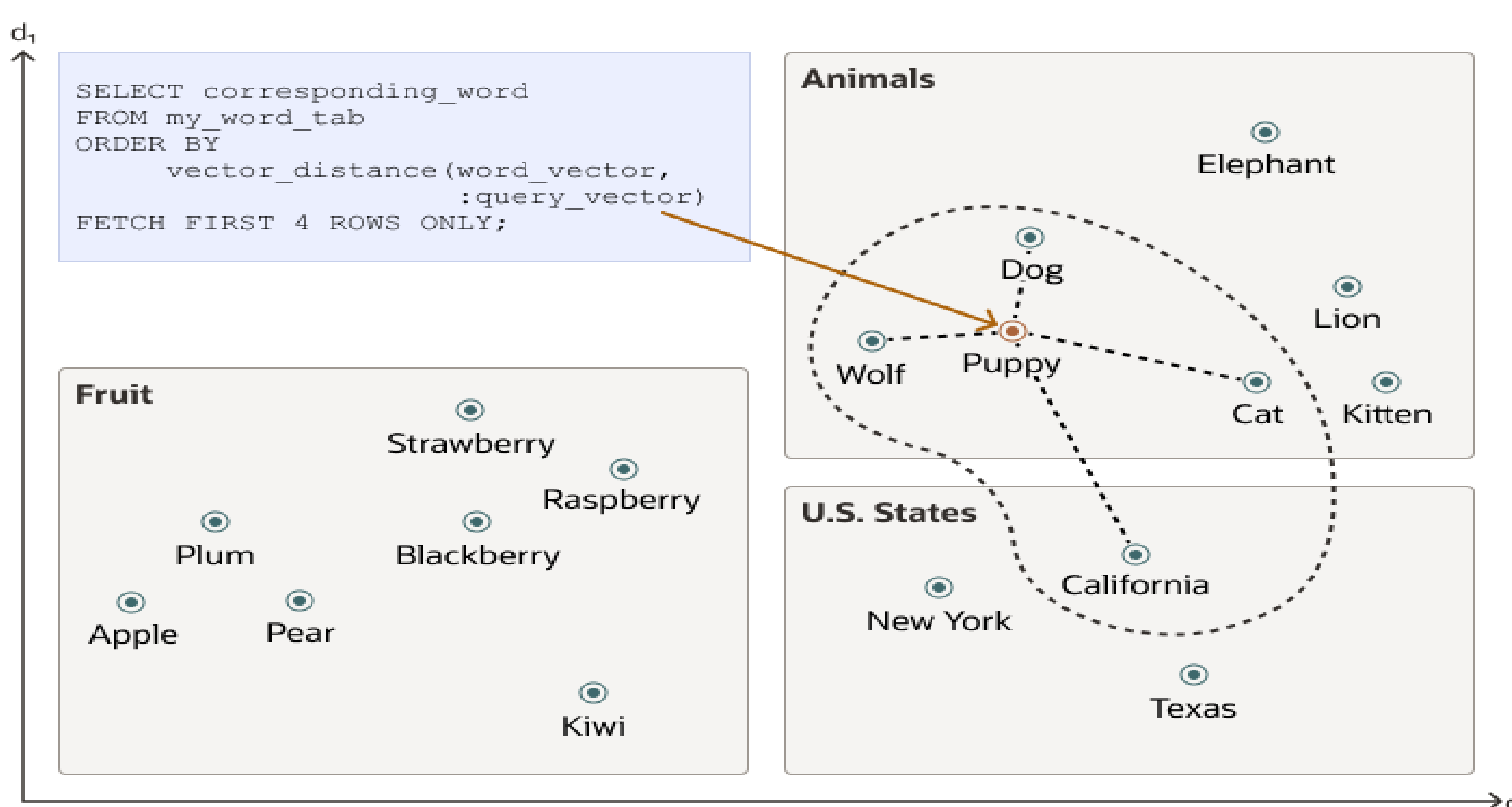
This project shows how vector databases enable fast context search over unstructured data using AI-generated embeddings. By storing and indexing high-dimensional vectors, these databases power smarter applications like recommendation systems, image search, and Natural Language Processing (NLP).

Vectors Embeddings

This simplified diagram illustrates a vector space where words are encoded as 2-dimensional vectors.



Similarity Search



Terminology

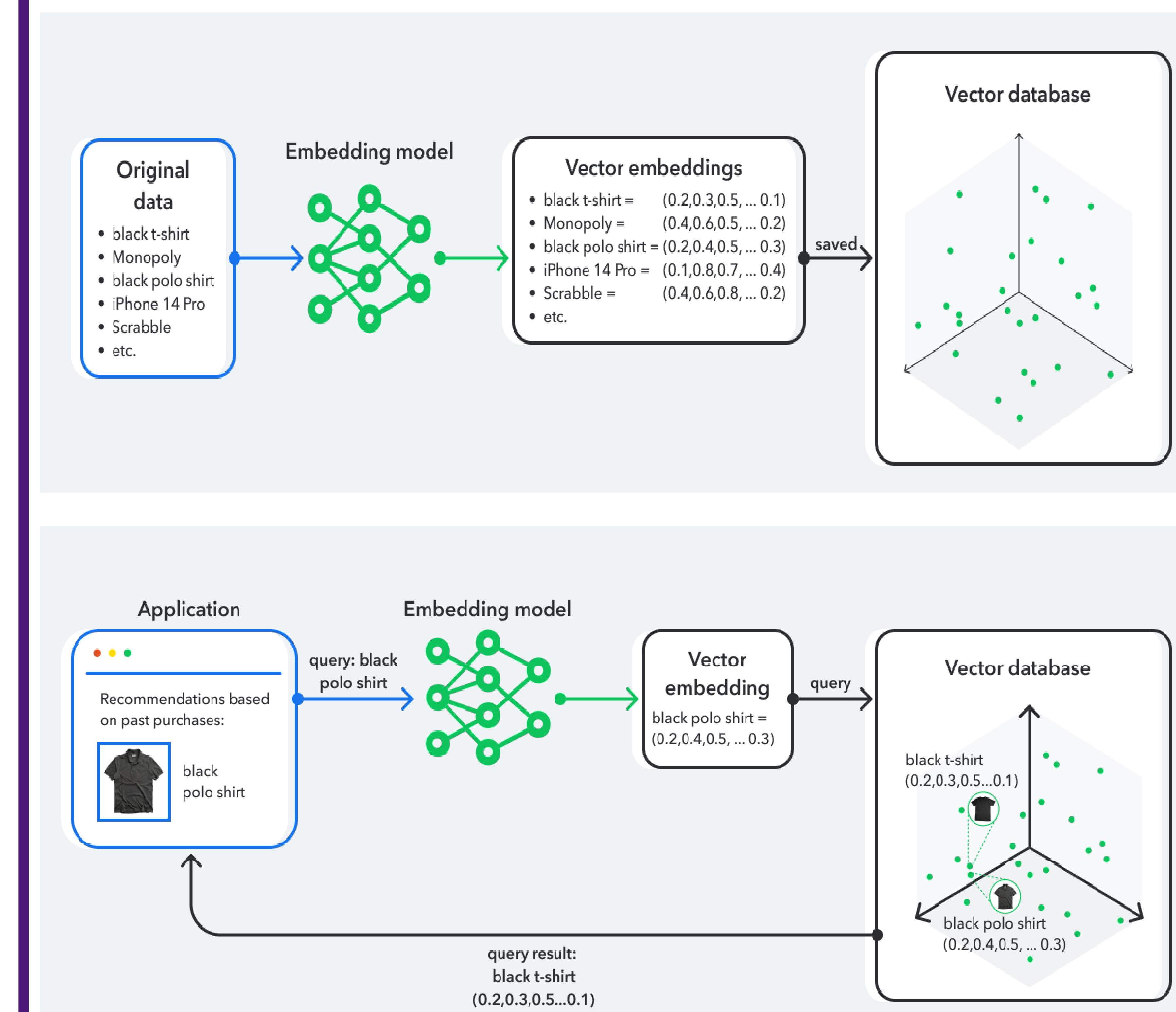
Vector Embedding: A numerical representation of unstructured data (text, image, etc.) used to capture semantic meaning.

Vector Database: A database optimized for storing, indexing, and querying high-dimensional vector embeddings.

Euclidian Distance: A metric that calculates the straight-line distance between two points (vectors) in space.

Embedding Model: A machine learning model (e.g., BERT, CLIP, GPT) used to generate vector embeddings from unstructured input.

Workflow



Technologies Used



Acknowledgement

A huge thank you to **Dr. Mei** and the **SERC Grant** who made this project possible.