AUTOMATING DIVERSE FILE EMBEDDING IN LANGCHAIN

IMPLEMENTING INCREMENTAL VECTOR STORAGE IN FAISS

HTTPS://GITHUB.COM/INSIGHTBUILDER

THE GOALS 🎯

- EMBED TEXT, PDF, CSV, EXCEL, CODE, IPYNB FILES
- INCREMENTALLY LOAD THE EMBED VECTORS INTO FAISS INDEX

STEPS TO ACHIEVE

REALIZE DATA IS STORED AS TEXT

LEARN HOW FILE LOADERS WORK

EXTRACT TEXT AND THEN SPLIT THEM

FEED THEM TO EMBED FUNCTIONS

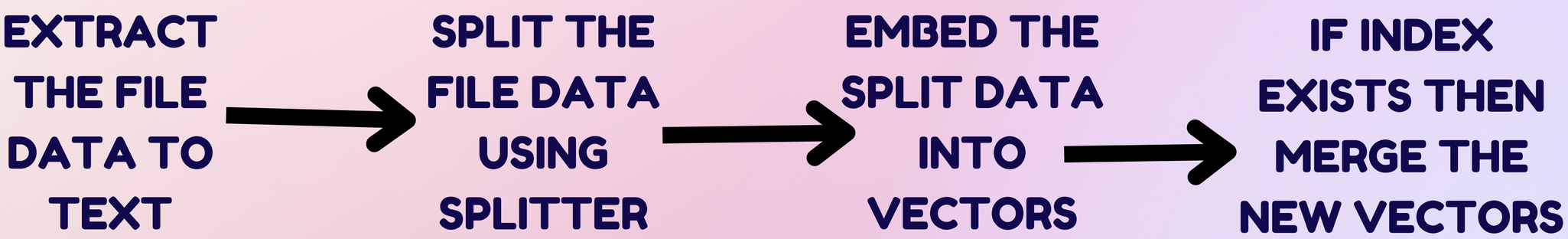
INDEX IS A DS TO WHICH NEW DATA CAN BE APPENDED

MODULARIZE THE CODE INTO FUNCTIONS

WRITE A SIMPLE TEST FUNCTION TO CHECK INGESTION

SEARCH THE VECTOR STORE & VERIFY

FLOW CHART



FAISS : FACEBOOK AI SIMILARITY SEARCH
OTHER INDEXES
CHROMADB
PGVECTOR
QDRANT
ANNOY

HTTPS://GITHUB.COM/INSIGHTBUILDER

TIME TO CODE

EACH OF THE FILE TYPES WILL POSE UNIQUE CHALLENGES WHEN IT IS GETTING EMBEDDED AND LATER RETRIEVED FROM THE INDEX. LETS LEARN THEM

HTTPS://GITHUB.COM/INSIGHTBUILDER

THANKS FOR WATCHING

PRACTICE

PRACTICE

LIKE

SHARE

SUBSCRIBE

PRACTICE

PRACTICE

HTTPS://GITHUB.COM/INSIGHTBUILDER