

Kirtankumar Parekh

Chicago, IL (Open to Relocate)

📞 224-251-9625 • ✉ kirtansp333@gmail.com • in kirtan-s-parekh • 🌐 bdk-333
<https://bdk-333.github.io/Kirtan-Parekh-Portfolio/>

Education

DePaul University

Master of Science in Data Science

Chicago, IL

Sep. 2024 – Jun. 2026

CVM University

Bachelor of Engineering in Computer Engineering

India

Jun. 2020 – Apr. 2024

Experience

Venom Technologies

Python Developer Intern

India

Jan. 2024 – Apr. 2024

- Designed and developed a full-stack Grocery Store Management System prototype, managing the end-to-end process from conceptual design to final implementation while emphasizing data management and integration.
- Engineered the backend using Python with the Flask framework and connected to a MySQL database, automating data handling processes and facilitating reliable information retrieval.
- Collaborated with a mentor through bi-weekly progress reviews, incorporating feedback to refine features and enhance overall system performance, demonstrating adaptability and effective team communication.

Projects

End-to-End Phishing Email Detector MLOps Pipeline

Python, DVC, MLflow, Docker, FastAPI, GCP Cloud Run, GitHub Actions

[Link](#)

- Engineered a reproducible binary classifier optimized for threat recall, achieving a 98.6% F2 score on a holdout evaluation of 16,496 unseen emails.
- Designed deterministic data pipelines using DVC and AWS S3, implementing custom validation stages to enforce schema consistency alongside MLflow for comprehensive experiment tracking.
- Automated CI/CD workflows with GitHub Actions for code testing and Continuous Machine Learning (CML) reporting, deploying the final model via a Dockerized FastAPI backend to GCP Cloud Run.

Cloud-Based Insurance Premium Predictor

Python, Apache Spark, PySpark, AWS EMR, Amazon SageMaker, Amazon S3, MLlib

[Link](#)

- Designed a scalable big data pipeline using Apache Spark on AWS EMR to process, clean, and perform complex feature engineering on 1.2 million synthetic insurance policy records stored in an Amazon S3 data lake.
- Leveraged Amazon SageMaker and Spark MLlib to train, tune, and cross-validate multiple regression algorithms, specifically addressing right-skewed financial target variables via log transformations.
- Uncovered highly non-linear relationships in insurance risk pricing, ultimately optimizing a Gradient Boosted Trees (GBT) regressor that significantly outperformed baseline linear models with a minimized RMSE of \$848.79.

Local RAG Document QA Engine

Python, Ollama (Llama 3.2), PyMuPDF, Sentence-Transformers, Vector Database

[Link](#)

- Designed a Retrieval-Augmented Generation (RAG) application enabling interactive, context-aware conversations with complex PDF documents, deploying local LLMs via Ollama to guarantee data privacy and security.
- Engineered an automated document ingestion pipeline featuring robust text extraction, semantic chunking, and high-dimensional vector embedding generation to optimize retrieval accuracy.
- Designed an inference engine that grounds AI responses strictly in retrieved document fragments, generating real-time source citations to significantly mitigate model hallucinations.

Technical Skills

Languages: Python, SQL, PySpark, R, Bash

Cloud & Big Data: AWS (S3, EMR, SageMaker), GCP (Cloud Run), Apache Spark

ML & AI: Scikit-learn, Spark MLlib, Hugging Face, Ollama, NLP / RAG

MLOps & Tools: Docker, MLflow, DVC, GitHub Actions (CI/CD), FastAPI, Git

Data & Viz: Pandas, NumPy, Matplotlib, Seaborn, Tableau, PyMuPDF