RISHIKESH DHAYARKAR

San Jose, CA | github.com | linkedin.com | medium.com | rishikeshdhayarkar1091@gmail.com | 347-471-8506

ML Engineer with 4 years of expertise in delivering production-grade machine learning solutions. Skilled in leveraging ML paradigms, LLMs, data & software engineering practices to develop products that solve complex NLP and AI challenges.

EDUCATION

NEW YORK UNIVERSITY — *New York City, NY*
Master of Science in Computer Science and Engineering — *Aug 2019 - May 2021*

VISVESVARAYA TECHNOLOGICAL UNIVERSITY — *Bangalore, KA*
Bachelor of Engineering in Electronics and Communication — *Aug 2014 - May 2018*

WORK EXPERIENCE

DATA SCIENTIST *(ML and NLP)* at FRACTAL — ***Sep 2021 – present | San Jose, CA***

- Developed an LLM-powered NLP application for rapid text analytics. Enabled UX researchers and analysts to extract insights from text data stored in csv files and SQL tables in minutes. Leveraged Python, Gemini models/APIs, TensorFlow, Flask, FastAPI, Hugging Face, and GCP to build the application. Implemented features including topic modeling, topic discovery, sentiment detection, text summarization, and question answering.
- Developed a comprehensive suite of LLM classifiers for identifying product issues, feedback, and concerns from users. (average F1 score of 96%). Implemented fine-tuned Gemini classifiers using Low Rank Adaptation (LoRA) approach, and developed few-shot classifiers leveraging advanced prompt engineering methods such as prompt chaining and chain of thought. The project utilized Python, Gemini models/APIs, and Hugging Face libraries, active learning tools to construct a robust labeled datasets for fine-tuning.
- Developed classification models for a leading video sharing platform, resulting in a 14% increase in Revenue over Ad Spend (ROAS) for the marketing team. Achieved this by constructing a customer sign-up propensity model (ROC AUC = 0.98, precision-recall AUC = 0.97) and a subsequent spend tier prediction model. Leveraged Python, TensorFlow, TFX, and BigQuery for model building, deploying, and setting up ETL pipelines for distributed data ingestion, cleaning, and preprocessing.

DATA SCIENTIST *(NLP)* at PROMAZO — ***Aug 2021 – Oct 2021 | Chicago, IL***

- Designed a similarity search-based model for a start-up, resulting in a 40% improvement in matching accuracy for their consumer product connecting students with industry experts for personalized career mentorship. Leveraged Python, Pytorch, Pandas, AWS, and HuggingFace, for feature extraction, transformer based embedding generation and model building and deployment.

RECENT PROJECTS (All Projects)

PODPROBE: PODCAST EXPLORATION TOOL | Github | *OpenAI API, Pinecone, Neo4j, LlamaIndex* — ***Jul 2024 – Present***

- Developed a dual Retrieval-Augmented Generation (RAG) system from scratch. Combined traditional RAG and graph-RAG to optimize rapid fact retrieval and complex entity mapping allowing users to quickly assess relevance and extract information from podcasts.
- Integrated the dual RAG system into a ReAct(reason and act) agent with memory, enabling efficient handling of complex multi-hop queries. This design minimizes calls to the RAG system, reduces latency, offers up-to-date information aligned with current trends.
- Leveraged OpenAI API for audio transcription and embeddings, Neo4j for constructing a knowledge graph, Pinecone for vector-based text storage, and LlamaIndex workflows for ReAct agent development.

ML-OPS PIPELINE FOR HATE SPEECH DETECTION | Github | *Pytorch, Docker, FastAPI, Google Cloud Platform* — ***Apr 2024 – Jul 2024***

- Built a production-grade MLOps pipeline for classifying hate speech, toxic comments, and cyberbullying on online platforms.
- Leveraged Docker for reproducibility, Dask for distributed data processing, and PyTorch Lightning for distributed model training, with deployment on Google Cloud Platform (GCP).
- Developed a scalable, maintainable, and cloud-ready infrastructure, serving the model via a REST API using FastAPI and Streamlit.

ML MODEL API FOR HOUSE PRICE PREDICTION | Github | *Python, FastAPI, Docker, Railway, Circle CI, Scikit-learn* — ***Feb 2024 – Apr 2024***

- Developed a house price prediction API using FastAPI, on a regression model trained on Kaggle's advanced house pricing dataset.
- Implemented end-to-end ML ops pipeline, including model packaging, containerization with Docker, and deployment on both PaaS (Railway) and IaaS (AWS EC2) platforms.
- Established CI/CD practices using CircleCI ensuring code quality, cross-environment compatibility through automated testing with Tox, and package distribution via Gemfury.

SKILLS

PROGRAMMING : Python, Java, Javascript, SQL, HTML, CSS, R
ML-OPS AND DEV TOOLS : Docker, Git, Weights & Biases, MLFlow, Circle CI, Github actions, TorchServe, Flask, FastAPI, Google Cloud Platform, AWS, Spark MLLib, PySpark, SparkSQL, BigQuery, Neo4j, Pinecone, DVC, Hydra-core
ML FRAMEWORKS AND TOOLS : Pytorch, TensorFlow, HuggingFace, OpenAI API, Gemini API, LlamaIndex, Langchain, Scikit-learn, Pandas, OpenCV, Nltk, Numpy, Gensim, XGBoost, LightGBM
ANALYTICS : Hypothesis Testing, Time-series Analysis, Causal Impact Analysis, A/B Testing, Statistical Inference, Excel, Statsmodels