Dante Secada-Oz

Email
contact@DanteOz.com

Website
DanteOz.com

Project
FastSearch

TL;DR

Experienced in building and deploying data/machine learning driven applications using real world data. Looking for a full-time position as a **Machine Learning Engineer** in NLP, information retrieval, recommender systems and LLMs on domain specific corpora.

EDUCATION

Rochester Institute of Technology '22 – B.S. Software Engineering / Math, RIT ▶

EMPLOYMENT

Two years of engineering co-op at Collibra ▶ – $5-billion enterprise intelligence and governance company listed ▶ as the 7th most valuable data startup in the world.

Machine Learning Engineer (Knowledge Graph) 2020 – Conducted tabular representation-learning research for entity deduplication, with all data remaining on edge for client privacy and security. Also, for Collibra-partner UC San Diego BlockLAB, worked on data mining research on whether corporate investment in data leads to significant ROI: implemented an ETL pipeline, transforming unstructured corpora of academic and business journals into a knowledge graph, using active learning, clustering, and topic modeling. Details ▶

Machine Learning Engineer (Business Process Automation) 2021 – Created and deployed a data pipeline to aggregate issues from engineering (Jira) and customer (Aha!) backlogs, and contextualize issues using customer and product metadata from various sources (Salesforce, Confluence, GitHub, etc). Redesigned ticket/feedback forms for ease of feature extraction. Created a classification model to automate allocation and prioritization of issues. Details ▶

PROJECTS : OPEN SOURCE

FastSearch – Built an end-to-end semantic search engine to help tens of thousands of students search the popular fast.ai ▶ ~300-hour machine learning video corpus. Performs low-latency retrieval and ranking of lecture transcripts (ONNX), with bi- and cross-encoder models trained using cross-architecture knowledge distillation (PyTorch), on a custom dataset containing ~1,000 fast.ai questions and ~27,000 lecture segments. Backed by a cloud data pipeline (Dagster) which scrapes and transcribes new video lectures (OpenAI Whisper) and incrementally updates an ANN search index (Qdrant). Tracks user queries and result feedback for model retraining. Deployed with fully custom CI/CD and MLOps (GitHub Actions) pipeline using IAC best practices (AWS CDK). MLOps launches backfill over the embedding/indexing pipeline and redeploys backend container with updated model weights upon push to model registry (Hugging Face). FastSearch website ▶   Project writeup ▶

PROJECTS : STARTUP

VizYourGov – A data-driven platform for visualizing the influence of money in U.S. politics, coming in 2024. Previous data pipeline was slow, monolithic, difficult to contribute to, had complicated inter-step dependencies and required constant manual monitoring. I rewrote the entire pipeline, consisting of 20,000 lines of Python (Dagster), 105 stored procedures, and 140 tables/views (dbt). I implemented structured logging, data catalog/lineage, data quality monitoring, end-to-end type safety (Dagster, Arrow ADBC, dbt) and async web scraping (HTTPX, Playwright). This resulted in a lower barrier to contribution, 25X faster scraping, asset level execution/scheduling, incremental partition materializations and end-to-end pipeline observability. Details ▶

PROJECTS : UNIVERSITY

Lockheed Martin – IoT Pipeline and Anomaly Detection (Graduation project) – Team built a desktop/mobile dashboard to monitor health and performance of factory machinery at the world's largest military and aerospace company. I built the ingestion/ETL pipeline for streaming data from factory floor, and a time series anomaly detection model for prediction and notification of machine failures, using AWS tooling for pipeline (Kinesis, SQS, Greengrass, MTConnect), hosting (Amplify), serverless backend (Lambda, APIGateway), NoSQL databases (DynamoDB, Neptune) and IAC (AWS CDK). Details ▶

JetBrains Research – Model deployment (SWEN Research) – Optimized model inference for AntiCopyPaster, a JetBrains IntelliJ IDE plugin, which detects and suggests Java code fragments to be refactored into methods on copy and paste. Prototyped Random Forest and XGBoost models to classify method re-factorability using code quality metrics such as cyclomatic complexity and per line connectivity. Productionized and deployed Random Forest (scikit-learn) and CNN models (Keras) in Java (ONNX Runtime). Details ▶

RIT

For summaries of a dozen SE/ML projects completed at RIT, please visit DanteOz.com ▶

Samples: Won sentiment analysis competition among graduate students (PyTorch) • Stock forecasting via sentiment analysis (PyTorch, XGBoost) • Stock sentiment analysis dashboard for Twitter (PyTorch, AWS: Lambda, DynamoDB, Amplify, Terraform, API Gateway) • 2D map routing via graph search (Java) • Spotify DBMS and analytics (Python, Postgres) • Stock portfolio management system design (Python, JavaScript) • Website fuzzer (Python, SQL, Beautiful Soup, Selenium) • Concurrent systems correctness verification for deadlocks and race conditions (Promela/SPIN). All RIT Projects ▶