KELVIN CHRISTIAN

Data Scientist & Machine Learning Engineer

206-747-9126 christian.ke@northeastern.edu kelvinc123 kelvinc123 kelvinc123.com

EXPERIENCE

Backend & ML Engineer (3-months contract) Feb 2024 – Present
Chekable *San Jose, CA*

- Developed multiple backend features using Flask and gRPC based on stakeholder requirements and business needs
- Deployed and managed 6 microservices seamlessly on Google Cloud Platform using Docker and Cloud Run
- Experimented with multimodal large language models (LLMs) such as Gemini and GPT for image-to-description tasks
- Developed a pipeline for patent generation by leveraging tools such as LangChain, Faiss, and GPT-4 with prompt engineering

NLP Research Intern Jun 2023 – Aug 2023
Institute for Experiential AI *San Jose, CA*

- Managed 50 parallel data processing jobs to process 6TB of LaTeX and PDF documents using Slurm, Python, and Bash
- Retrieved 1.7M+ table images and text from PDFs using the Table Transformer and OCR models from Hugging Face
- Implemented a model training pipeline in PyTorch to learn the table representation mapping from table image to LaTeX code

Data Scientist Intern Jul 2022 – Sep 2022
Cisco Systems *Seattle, WA*

- Conducted data cleaning and analysis on a 1.2TB real-time web query dataset from AWS S3 with PySpark to improve DNS-layer network security
- Reduced ETL time on Databricks by 70% through the optimization of PySpark SQL query and data migration to Delta Lake
- Enhanced user experience by cutting 6% of incorrect website blocks through NLP feature engineering on domain names
- Initiated an anomaly detection system based on users' request patterns by clustering 100,000+ users into 60 profile groups

AI Engineer Intern Jan 2022 – Jun 2022
Aeyesafe *Seattle, WA*

- Developed a microservices-based ETL pipeline on AWS to process real-time sensor data by using RabbitMQ and gRPC
- Utilized Spark and Airflow to automate the generation of weekly business intelligence reports from large-scale datasets
- Wrote 15+ Flask RESTful APIs containing complex SQL queries to manage user profiles, device data, and health metrics
- Led a team of 4 interns in the research and development of diverse machine learning models for fall detection
- Trained and deployed a computer vision model to predict human posture from the thermal camera with 96% test accuracy

PROJECTS

iSEArch Hotels | *Python, Docker, Streamlit, Qdrant, LangChain, OpenAI* March 2024

- Built and deployed a RAG hotel search chatbot by leveraging the Qdrant vector database as a recommendation system
- Won 2nd place in the Traversaal.ai Hackathon

Movie-style Art Generation | *Python, PyTorch, OpenCV, Hugging Face* Dec 2023

- Implemented a pipeline to enhance dataset quality by applying super-resolution on movie frames using Real-ESRGAN
- Fine-tuned Stable Diffusion model using LoRA and Dreambooth methods, generating high-quality movie-style artwork

Tennis Action Recognition | *Python, PyTorch, Hugging Face, Tensorboard* Jun 2023

- Fine-tuned the TimeSformer model on the tennis clip dataset and achieved classification test accuracy of 91%

EDUCATION

Northeastern University San Jose, CA
M.S. Computer Science, GPA: 3.88/4.00 *2021 – 2023*

University of California, Los Angeles Los Angeles, CA
B.S. Financial Actuarial Mathematics (Minor in Statistics), GPA: 3.58/4.00 *2017 – 2019*

SKILLS

Machine Learning: PyTorch, Tensorboard, Spark, OpenCV, NLTK, Scikit-learn, Pandas, Statsmodels, NumPy
Programming Languages: Python, R, SQL, C++, Java, Javascript
Tools: AWS (EC2, RDS, S3, ECR, Lambda), GCP (GCE, Vertex, Cloud Run), Docker, Grafana, OpenAI, LangChain
Databases: MySQL, MongoDB, Redis, PostgreSQL, Qdrant, Milvus, Weaviate, Elasticsearch, InfluxDB