

Dan Huynh

☎ 226-606-9284 | ✉ dan.huynh@uwaterloo.ca | 🔗 LinkedIn | 🐙 GitHub | 📁 Portfolio

OVERVIEW

Languages: C/C++, Java/Scala, Python, TypeScript, GO, PHP, SQL, Swift, MATLAB, VHDL

Technologies: Docker, Azure, GCP, AWS, Node, Spring, React.js, Angular, Scikit-learn, Tensorflow, ROS2, OpenCV

5+ years of experience in designing mechanical equipment using SolidWorks and the Autodesk suite

PROFESSIONAL EXPERIENCE

Data Scientist @ theScore/ESPN Bet

Sept 2024 – Present

- Conducted an **EDA** using **pandas** on proprietary stolen base odds generated from an **sklearn** logistic regression model trained on SportRadar data, identifying a systematic under-prediction for explosive base runners and reducing deviation from consensus odds by 56%.
- Developed and deployed an automated **Python**-based data monitor to validate projection data availability and sparsity across 27 tables in **BigQuery** and **Postgres**, triggering real-time **Slack API** alerts to an on-call channel for immediate pipeline remediation via an **Argo Workflow**.
- Currently building infrastructure to enhance machine learning pipeline robustness and improve MLB projections.

Software Development Engineer in Test @ Vivid Seats

Jan 2024 – Apr 2024

- Modified a **Selenium** grid interface to display repository-specific metadata, leading to a 100% increase in test identification across 71+ regression suites.
- Designed a **Stoptlight**-documented test-data Backend for Frontend (BFF) with **Spring** and **OpenAPI**, exposing RESTful endpoints for **JPA** entity generation and management that insert regression-agnostic data for E2E checkout tests into 4+ Vault-authenticated **AWS Aurora** databases.
- Developed an embedded subscriber for the test-data BFF which enables on-demand and autonomous data cleanup, via client API calls and a proprietary cleanup micro-service, allowing for concurrent data management.
- Integrated **SonarQube** for static code analysis, leading to the development of 210+ unit and integration tests using **JUnit5**, **Mockito**, and **Spring** that achieved 96% code coverage for the test-data BFF.
- Collaborated with 5+ teams to troubleshoot and resolve Jenkins build issues across 30+ repositories, contributing to CI/CD pipeline stability by reducing build failures.

Data Scientist @ PureFacts Financial Solutions

May 2023 – Aug 2023

- Developed and tuned an **Scikit-learn** Bayesian optimized random forest regressor with a mean percentage error of 17.32% that forecasts client revenue movements, whilst providing interpretable explanations for model predictions using **SHAP**.
- Aggregated, cleaned, and wrangled over 100,000 rows of data using **pandas**, and performed missing data and outlier treatment, resulting in a 94.86% decrease in the mean percentage error of a random forest regressor.
- Designed a dashboard using **Plotly Dash** that features dynamic visualizations of investor revenue, AUM, transactions, and customer trends over time for PureFacts clients, encouraging data-driven decision making.
- Led development of a **Flask + React** tool tailored to the PureFacts tech stack utilizing **OpenAI APIs** that empower non-technical personnel with accessible information and optimizes engineer labor time whilst maintaining data confidentiality.

Software Engineer @ Ford Motor Company

Sept 2022 – Dec 2022

- Created components including a data-model agnostic autocomplete component using **React Typescript**, that queries 1000+ **Firestore** records for objects that fit a Regex string on one of 7+ record properties.
- Wrote asynchronous RESTful methods using **Axios** that reads/writes to 1000+ records in a **Firestore** database.

PROJECTS

Perceptions Lead @ Watonomous — LiDar Object Detection | Github

Sept 2023 – Sept 2024

- Developed a proprietary data loader for **OpenPCDet** to processes 32-beam, 4/5 feature Velodyne point clouds into **NumPy** arrays, optimized for **VoxelNeXt**, **TransFusion**, and **PV-RCNN** predictions.
- Wrapped **OpenPCDet** in a **ROS2-humble** node that processes a point cloud rosbag feed, publishing real-time bounding box predictions through the **Foxglove** WebSocket protocol for immediate data visualization.
- Designed a utility that converts **PyTorch** tensors into PointCloud2 messages, post-processed with dual transformation matrices that project **LiDar** PointCloud2 data onto bounding-box prediction planes.
- Modified **OpenPCDet** visualization utilities to render static **PV-RCNN** bounding-box predictions using **XVFB**, ensuring compatibility without reliance on a native **X-11 server**.

EDUCATION

University of Waterloo — GPA: 3.9/4.0

Waterloo, Canada

Honours B.ASc Candidate in Mechatronics Engineering (Option in Software Engineering)

2021 – 2026