

Summary

- Interested in building Machine Learning and Data products and infrastructure at scale.
- Work experience in Machine Learning infrastructure particularly on automatic retraining pipelines.
- Research experience on live data collection systems for data analysis.
- Contributed to several open-Source projects like Apache Airflow, Kubeflow Pipelines and Tensorflow Extended.

Experience

- **Twitter** *Senior Machine Learning Engineer, April 2022 - Present*
 - Technical lead for the ML Pipelines team on Cortex Platform.
 - Led KFP SDK internal general adoption across the company.
 - Worked on pipelines python client authentication and next-gen component authoring.
 - Built internal command line tool for KFP Pipelines continuous release.
 - Worked as point of contact for industry collaborations with Spotify and Google.
 - Mentored team members helping scope and drive projects for the ML Pipelines team.
 - Defined quarterly roadmap, drove project alignment across teams, improved team execution with Kanban.
- **Twitter** *Machine Learning Engineer II, October 2020 - April 2022*
 - Led development for team CI/CD system improving code and documentation stability.
 - Developed initial python wheel and VSCode Twitter monorepo integrations used by 100s of engineers.
 - Co-led/designed TFX pipeline production scheduling for Kubeflow Pipeline clusters.
 - Led pipeline integration for GPU training using AI Platform Training.
 - Co-authored evaluation for next-gen ML Training Service on GCP.
 - Taught "ML at Scale on GCP with TFX" class for internal Hackweek with over 300 virtual attendees.
 - Collaborated with customer teams to launch the first ads click model migration to GCP with improved developer iteration, leading to faster model improvements.
 - Cross-team mentoring to build new pipeline components that integrate well across platform.
 - Member of various working groups like ML to GCP virtual team, and Python Customer Advisory Board.
 - Multiple contributions to TFX, and Kubeflow Pipelines OSS projects.
- **Twitter** *Machine Learning Engineer, September 2019 - October 2020*
 - Contributed to internal distributed Bayesian hyperparameter optimization framework based on Ax.dev.
 - Supported ML research work for "Tuning Word2vec for Large Scale Recommendation Systems" (RecSys2020).
 - Worked on evaluation of industry pipeline solutions and eventual adoption of TFX for defining end-to-end ML pipelines.
- **Project EPIC, University of Colorado Boulder** *Research Assistant, February 2017 - August 2019*
 - Analyzed 2000 tweets/minute of real time data by building Big Data pipeline using Spark, Kafka and Cassandra.
 - Deployed system prototype with Kubernetes using custom microservices written in Go and Python.
- **Twitter** *Software Engineering intern, May 2018 - August 2018*
 - Developed CLI to deploy Apache Airflow instances internally.
 - Helped reduce ML model training and deploy time by 1/4 developing ML Workflows.
- **inLab FIB** *Web Development Intern, March 2015 - January 2017*
 - Built REST API for faculty data - Average usage of 25.000 daily accesses. (Django, Python)
 - Reduced Oracle SQL query response time by 1/3 using non-blocking updates and materialized views.
- **con terra GmbH** *Intern, Summer 2015*
 - Developed bike share routing app using OpenStreetMap open data and Citibikes API. (Javascript)

Education

- **University of Colorado Boulder** Boulder, CO
M.S., Computer Science (GPA 3.97) *August 2017 - August 2019*
 - **Courses:** Analysis and Design of Algorithms, User Centered Design, Database Systems, Big Data Architecture, Machine Learning, Computational Genomics, Natural Language Processing.
- **Universitat Politècnica de Catalunya** Barcelona, Spain
B.S., Computer Science (8.8/10) *September 2013 - July 2017*
 - **Courses:** Data Mining, Data Compression, Programming Languages, Software architecture, Databases, Parallelism, Project Management, Web architecture, Android, NoSQL databases, Reflective programming.

Open Source projects

- **TFX Addons SIG** *Core Contributor, May 2021 - Present*
 - Release manager from version 0.1.0 to date.
 - Mentored new contributors developing new components with reliable tests, and consistent codebase practices.
 - Contributor and maintainer for project CI/CD system, XGBoostEvaluator, FeastExampleGen and others.
 - Collaborator for ModelCardTooling migration mentoring Twitter team members to contribute to project.
- **AIP-31: Airflow functional DAG definition / "Taskflow API"** *Core Contributor, January 2020 - July 2020*
 - Proposed and implemented improvement for Airflow DAG user experience adding explicit XCom data.
 - Released in Airflow 2.0.0 and used in one of the main tutorials for the project.

Awards

- **Google OSS Peer Bonus for contributions to TFX and TFX Addons** (2022)
- **Denver Government award winner at ETHDenver.** (2018)
- **Grant winner at MD5 Physical Cyber Hackathon.** (2018)
- **Balsells Fellowship.** (2017-2019, two years of funding for graduate studies)
- **Balsells Mobility program.** (2017, five months of funding for research)

Publications

- Jennings A., Gerard C., Kenneth M. A., Leysia P., and Rebecca M. **Incorporating Context and Location Into Social Media Analysis: A Scalable, Cloud-Based Approach for More Powerful Data Science.** *HICSS 52 (accepted).*
- Gerard C. **Social Media Analysis for Crisis Informatics in the Cloud** *MS Thesis*

Skills

- **Programming:** Java, Python **Experience with:** Kubeflow Pipelines, Tensorflow Extended, Django/Flask/FastAPI, SQL, Scala, Tensorflow, Apache Airflow, Apache Beam/Dataflow, HTML, CSS, Haskell, Smalltalk, Go, Spark, Kafka, Kubernetes
- **Native:** Catalan, Spanish. **Advanced:** English.
- **Hobbies:** Skiing, Running, Photography, Roller hockey, Piano, Swim.