Ryan Johnson

Data Scientist *Seattle, WA*

+1 (253) 315-0389 | riosanjuan314@gmail.com | rio-sanjuan | rtjohnson12

Skills

Programming Languages	Python, R, Scala, SQL
Machine Learning	Supervised Learning, Clustering, A/B Testing, Ensemble Learning, Dimensionality Reduction, Time-Series Analysis, Feature Engineering & Selection, Deep Learning, Graph-Based Learning
MLOps	AWS, Docker, Kubernetes, Airflow, Terraform, Jenkins, Git

Work Experience

Data Scientist

Jan 2022 - Present

Chewy

Bellevue, WA

- Designed and deployed an end-to-end ML pipeline for forecasting Autoship orders. Combining subscription-level features with PySpark and AWS Glue, training and inference with AWS Sagemaker, and orchestration through Airflow dags, the pipeline involves automated ETL, tuning, and publishing, outperforming existing heuristic-based approach by up to 5% MAPE at the weekly level.
- Designed and deployed a Hierarchical Reconciliation framework as a callable AWS Sagemaker processing job, allowing for seamless integration with existing forecasting systems and providing a robust and scalable solution for reconciling forecasts across multiple granularities. This framework is now used by a majority of our production ML pipelines.
- Designed and deployed a forecast evaluation framework, allowing for automated comparison of forecast performance across our product catalogue, enabling us to rank model variants for specific use-cases like holiday planning, short-range labor planning, and long-lead buying decisions.
- Led multiple white-paper reviews within the team covering the latest research in both the field of Data Science and the eCommerce/operation planning space.

Data Scientist

Sep 2019 - Jan 2022

The Energy Authority

Bellevue, WA

- Overhauled Hydro-Power Forecasting pipeline by integrating new data feeds from NW river system and plant-level generation. The system ingests and cleans data from thousands of river meters and achieved a 5% reduction in system-level forecast MAPE.
- Designed and lead transition to new CI/CD system for running containerized models with Docker and GitHub Actions, enabling streamlined deployments and establishing new team standards for project version control and production deployment.
- Worked with Software Engineering team to develop model failsafe system. In the case of critical data feed failures, production models revert to failsafe mode and utilize a reduced set of features, ensuring that our real-time traders will always have actionable forecasts 24/7.

Education

MS in Applied Mathematics

Dec 2019

University of Washington

Seattle, WA

- Professional Excellence Award for distinguished academic merit and professionalism

BS in Mathematics and Computer Science

Jun 2018

Santa Clara University

Santa Clara, CA

- Paul R. Halmos Award for Outstanding Academic Record
- President of Pi Mu Epsilon, Mathematics Honor Society