

Julien J. Pierret, Ph.D.

Olympia, WA | julien.pierret@gmail.com | GitHub: <https://github.com/dafrenchyman>

Applied AI / Machine Learning Engineer

Applied AI and Machine Learning Engineer with a Ph.D. in Computational Statistics and deep experience building production AI systems, LLM-powered workflows, ML platforms, and distributed data pipelines. Strong background combining machine learning, statistical modeling, software engineering, cloud infrastructure, and AI product development. Experienced turning research prototypes into production features and building practical AI systems for fraud, AML, document understanding, NLP, and workflow automation.

CORE COMPETENCIES

Applied AI / ML: LLM Applications, AI Agents, NLP, BERT, Scikit-learn, Feature Engineering, Predictive Modeling, Fraud Detection, AML Modeling, Statistical Modeling

AI Systems / Data Infrastructure: PySpark, Databricks, Snowflake, Spark, Airflow, Kubernetes, Docker, Terraform, Pulumi, AWS, Helm, CI/CD

Languages / Platforms: Python, Java, C#, SQL, Linux, NixOS, nix-darwin, Bash

SELECTED PROJECTS & PUBLIC WORK

GitHub: <https://github.com/dafrenchyman>

Climate Visualization Platform: <https://icharm.sdsu.edu> - Interactive large-scale climate data visualization platform developed as part of Ph.D. research and still being updated

Machine Learning Engineering Course Materials: <http://teaching.mrsharky.com> - Public course materials developed while teaching Machine Learning Engineering at San Diego State University.

PROFESSIONAL EXPERIENCE

Senior Machine Learning Engineer

2021-Present

Unit21

- Built AI agent systems and LLM-powered workflow automation tools for financial investigation, data analysis, and customer-defined task execution.
- Developed AI systems capable of processing user-defined analytical tasks involving structured datasets and image-based inputs.
- Led AI/ML research prototypes that informed or evolved into production-facing platform features and products.
- Designed and built ML systems for AML and fraud detection using heterogeneous, customer-specific datasets across hundreds of customer environments.
- Built automated feature generation pipelines for customer-provided custom data, enabling substantial model performance gains for customers with rich, predictive data sources.
- Improved model performance in high-lift customer environments from approximately 0.60 AUC to as high as 0.93 AUC through automated feature engineering and model pipeline improvements.
- Researched consortium-based fraud intelligence sharing approaches to improve cross-customer risk detection and bad-actor identification.
- Re-architected large-scale Spark processing from Spark SQL to optimized RDD-based workflows, reducing runtime from approximately 12 hours to roughly 30 minutes on the same cluster.
- Migrated ML processing workflows from Snowflake-based execution to PySpark/Databricks.
- Designed and maintained Nix/nix-darwin infrastructure for macOS engineering environments, standardizing dependencies and improving reproducibility across engineering teams.

Adjunct Professor - Machine Learning Engineering

2020-2023

San Diego State University

- Developed and taught Machine Learning Engineering coursework focused on practical ML systems, model deployment, production workflows, and applied engineering patterns.
- Created public lectures, slides, and examples used by students learning modern ML engineering concepts.

Senior Machine Learning Engineer

2018-2021

AppFolio

- Built conversational AI and customer-support chatbot systems using BERT-based NLP models and custom feature engineering.

- Developed document understanding models to extract invoice metadata including vendor, invoice amount, invoice date, and related fields.
- Built a fully automated ML training and deployment platform supporting approximately 1,400 customer-specific invoice prediction models.
- Designed end-to-end retraining workflows that trained, validated, packaged, and deployed individualized customer models into production.
- Implemented CI/CD workflows for automating data preparation, model training, validation, deployment, and production updates.
- Designed cloud-native ML infrastructure using Kubernetes, Docker, Airflow, Terraform, and production deployment workflows.

Data Scientist

2014-2018

Experian

- Built large-scale model optimization systems on AWS EMR to evaluate candidate variables and identify high-performing feature combinations.
- Developed distributed analytical systems using Hadoop, Pig, Java, SQL, Spark, and Airflow for feature analysis, model evaluation, and reporting.
- Designed optimization workflows for evaluating large combinatorial modeling spaces that were computationally infeasible to search exhaustively.
- Re-architected Hadoop-based data processing workflows to Apache Spark, reducing processing time and AWS infrastructure cost.
- Created systems to analyze product logs and derive rules, logic, and predictive signals for transaction resolution workflows.
- Expanded available predictive variables from approximately 100 to over 2,000 through automated feature engineering and data analysis.

Statistical Modeler

2012-2014

Digital Risk

- Built end-to-end predictive modeling systems using credit, property, DMV, web, API, and third-party data sources.
- Designed and maintained a SQL Server research database with over 9 million observations and more than 5,500 candidate predictive variables.
- Developed mortgage and financial risk models using R, SAS, Matlab, SQL Server, and C#.
- Implemented production scoring systems and automated customer reports containing maps, visualizations, charts, and analytical summaries.
- Built optimization systems and multithreaded analytical tooling for modeling and operational workflows.

.NET Application Developer / QA / Technical Support

2002-2012

CoreLogic

- Built fraud detection and predictive analytics systems for financial services clients.
- Developed survival analysis systems, net-present-value calculations, and optimization workflows for financial applications.
- Built a credit card bust-out fraud model for a major banking institution, covering data cleaning, feature extraction, modeling, and delivery.
- Re-engineered production mathematical optimization logic, improving system performance by more than 10x.

EDUCATION

Ph.D., Computational Statistics - San Diego State University / Claremont Graduate University, 2018

M.S., Statistics - San Diego State University, 2014

B.S., Statistics & B.A., Mathematics - San Diego State University, 2009 - Most Outstanding Math/Stats Graduating Student, Class of 2009

PUBLICATIONS

Pierret, J., and S.S.P. Shen, 2017. 4D visual delivery of big climate data: A fast web database application system. *Advances in Data Science and Adaptive Analysis*.

Shen, S.S.P., Pierret, J., et al., 2020. 4DVD visualization and delivery of 20th century reanalysis data: methods and examples. *Theoretical and Applied Climatology*.