

AUSTIN WINDSOR

*Data Scientist /
Analytics Engineer /
Applied ML*

📍 Taipei

EDUCATION

Masters of Science
Computer Science (AI/ML)
Georgia Institute of Technology

Bachelor of Science
Industrial Engineering
(Operations Research)
Georgia Institute of Technology

CERTIFICATIONS

- AWS Machine Learning, AWS Developer, AWS Solution Architect

SKILLS

Machine Learning

- Causal Inference, Clustering, A/B Testing, Predictive Modeling, Recommender Systems, Churn Modeling, Customer Segmentation, NLP, Regression, Network Science, PyTorch

Data Engineering

- Relational Databases, ETL/ELT Pipelines, dbt, Semantic Layering, PySpark, Microservices, Data Quality Frameworks, PySpark

Programming

- Python (pandas, scipy, numpy), SQL, R, bash

Leadership

- Agile (Scrum Master), Jira, Stakeholder Management, Project Planning, Mentoring, FRQ Bidding

Maths

- Linear Optimization, Discrete Math, Calculus, Linear Algebra, Statistical Modelling, Simulation

Languages

- English (Native), Chinese (C1), French (B2), Spanish (B1)

CAREER OBJECTIVE

Data Scientist and Analytics Engineer with 7 years of experience building machine learning, experimentation, and cloud analytics systems across consumer media and healthcare. I specialize in recommender systems, causal/product analytics, data platform design, and production data workflows. My work focuses on turning ambiguous business problems into reliable models, metrics, pipelines, and decision tools.

WORK EXPERIENCE

Lead Data Scientist

PorticoMedia

📅 Jan 2023 - current 📍 Taipei, Taiwan

- Led end-to-end ML and analytics initiatives, translating product and marketing goals into production-grade data solutions across personalization, churn, and engagement.
- Designed and deployed recommender systems (Python, AWS SageMaker) using user-item and temporal signals, increasing CTR by 3%.
- Built churn prediction and LTV models (scikit-learn, causal inference techniques), enabling targeted campaigns that reduced churn by ~5%.
- Architected and maintained 10+ AWS microservices (Lambda, Step Functions, EventBridge) for real-time data processing and model orchestration.
- Developed scalable data pipelines (SQL, PySpark, Glue, RDS) processing 100M+ records, supporting analytics and ML feature generation.
- Implemented data quality validation frameworks using Python and AWS, resulting in a 30% increase in the reconciliation of third-party and internal data features.
- Built 8+ BI dashboards (QuickSight, Tableau) and self-service datasets, reducing ad hoc analysis requests and improving decision velocity.
- Modeled user behavior via clustering (DBSCAN, kNN) to segment viewing patterns and inform content and marketing strategies.
- Designed and standardized semantic data layers using dbt, improving data consistency and documentation across teams.
- Defined core business metrics (engagement, retention, conversion) and aligned stakeholders on measurement frameworks, boosting clarity in performance evaluation.
- Led cross-functional project scoping, prioritization, and delivery using Agile methodologies (Jira), ensuring timely execution of analytics initiatives.

Data Scientist Manager

KPMG

📅 June 2019 - Aug 2022 📍 Chicago, USA

- Led data science delivery for \$14M+ federal healthcare program, analyzing nationwide insurance and ICD9 claims data.
- Managed and scaled two teams (4 + 8 DS/analysts), driving execution across modeling, data engineering, and reporting workflows.
- Designed anomaly detection pipelines (Python, SQL), performing data validating for 50M+ records weekly, identifying ~12% policy violations.
- Built and deployed ETL pipelines (PySpark, AWS EC2/RDS) to process large-scale healthcare datasets and support compliance workflows.
- Developed statistical analyses on risk-adjustment policies using ICD-9 claims, premiums, and risk scores for federal reporting.
- Created 15+ dashboards (Tableau) to visualize anomalies and market dynamics for internal and government stakeholders.
- Acted as primary technical liaison across vendors and federal stakeholders, aligning requirements and resolving delivery risks.
- Translated research and policy documentation into production analytics frameworks with measurable business impact.
- Led Agile delivery, managing roadmap, prioritization, and technical blockers across 7 workstreams to decrease spillover by 30%.
- Authored successful RFQ proposals, contributing to \$10M+ contract expansion.

Natural Language Understanding Research Scientist

NTT Communication Labs

 May 2018 - Aug 2018  Kyoto, Japan

- Evaluated performance of deep learning RNN and attention models (PyTorch) on semantic benchmarks, uncovering dataset and architecture bottlenecks that informed model improvements.
- Augmented training corpora using word embeddings (word2vec) to improve semantic representation and model performance.
- Authored technical research reports to support the publication 'Can a Robot Get into the University of Tokyo?'