ANSHUMAN A. KUMAR

anshuman27kumar@gmail.com | 716-867-8146 | linkedin.com/in/anshuman27kumar

Cambridge, Massachusetts

Experienced Machine Learning professional with a background in deep learning and statistical modeling, exploring opportunities in foundation model building, post-training, and designing agent-based systems. I'm motivated by the potential to apply my skills to real-world problems that enhance human decision-making and empower individuals through AI.

Work Experience

Akamai Technologies

Principal Data Scientist **2024-Present**

- Owning ML and AI initiatives for product recommendations, customer behavior modeling, and marketing data science, to drive Akamai's Go-To-Market strategy, with a focus on cross-functional alignment
- Overseeing development of ML solutions end-to-end, from data ingestion to production deployment, while overseeing on-prem infrastructure, ML tooling, CI/CD workflows, and release management
- Leading a globally distributed team of 4 data scientists, providing technical guidance, managing team's project delivery & stakeholder alignment, and supporting the team's technical mentorship & career development

Key Project: Agentic workflows for Customer Journey Optimization & Insights Narration

- Conceptualizing and prototyping a multi-agent AI system, integrating generative models, reinforcement learning, and traditional ML to automate lead generation, contact enrichment, and personalized marketing creative generation
- Developing an AI-powered Insights Narrator, that translates complex SQL, BI reports, and business KPIs into real-time natural language insights for stakeholders, in an effort to reduce time-to-insight and minimize manual analytics
- Builing topic-level intent segments by summarizing content engagement across customer interactions with LLMs and generating embeddings for clustering and sentiment analysis, for refining product messaging

Key Project: ML and Behavioral Modeling to drive Product-Led Growth for Akamai's Cloud Business

- Mentored a data scientist in building a predictive lead scoring system using LightGBM based on 7-day usage behavior; with model achieving 78% recall in flagging high-MRR accounts ($500+) for Sales routing
- Led development of an unsupervised K-Medoids segmentation model incorporating product usage, firmographics, and RFM features; identified 9 actionable customer segments to drive personalized outreach and in-product engagement
- Clustered top-performing segments by usage sequence patterns and applied Heuristic Miner process mining algorithm to uncover adoption flows; insights are shaping gamification strategies for product engagement
- Mentored a junior data scientist in designing a causal inference framework (case-control analysis) using observational data to assess campaign effectiveness and deliver post-hoc ROI insights in the absense of A/B testing
- Led development of a Nested Multinomial Logit Discrete Choice Model in collaboration with two data scientists, using panel data to support product bundling strategies

Lead Data Scientist **2022-2024**

Key Project: New Customer Fit Modeling and B2B Product Recommendation Models

- Developed product recommendation models using a Positive-Unlabeled (PU) Learning framework trained on labeled product purchases and unlabeled prospects, achieving 74% average recall in identifying high-fit customers
- Designed a dual-encoder neural network in PyTorch for mixed-type tabular data, integrating categorical embeddings and numerical features, trained using a custom PU contrastive loss to learn discriminative representations for downstream classification
- Integrated Gaussian copula–based synthetic data generation with K-NN sampling to construct augmented minibatches for contrastive loss training, enabling representation learning under PU constraints and limited label availability
- Boosted Akamai's new customer acquisition by building a propensity-to-buy model using LightGBM with PU learning weights, uncovering $8.7M in sales pipeline and driving $6.8M in GMRR since 2023

Key Project: Journey-based Marketing Mix Modeling (MMM) for Channel optimization

- Implemented a novel MMM framework using Double Machine Learning (DML) to estimate causal-effects of different marketing channels (concurrent multi-valued treatments), within B2B journey stages
- Applied Principal Component Analysis (PCA) to handle cross-channel correlations, to generate uncorrelated shared marketing factors for causal estimation using DML
- Utilized DML's partial linear model, a multi-output random forest regression model, and an inverse PCA transform to estimate the causal effects of marketing channels with cross-channel interactions
- Implemented modular, production-grade ETL pipelines using Hamilton to orchestrate feature transformations for MMM, enabling clear dependency tracking, reproducibility, and scalable DAG-based execution

Senior Data Scientist **2017-2022**

Key Project: Multi-Stage Anomaly Detection & Root Cause Analysis for Customer Incident Prediction

- Developed a multi-staged anomaly detection system to proactively identify customer-impacting incidents triggered by software rollouts across Akamai's global network regions, based on sever and network telemetry data
- Used Locality Sensitive Hashing (LSH)-based ANN scoring on aggregated statistical features, to efficiently pre-filter anomalous time windows, triggering downstream systems for more fine-grained detection
- As part of the detection system, applied Isolation Forest on high-resolution metrics to identify point anomalies, and an LSTM-based Autoencoder trained on normal behavior to detect temporal shifts across historical windows using reconstruction score
- Built scalable PySpark pipelines to process unstructured telemetry data for end-to-end model training and deployment
- Built a contextual anomaly detection engine, that used clique detection algorithms to trace root causes of performance degradation issues, by modelling alert dependencies and correlated failure patterns as a contextual graph

Data Scientist **2015-2017**

Key Project: Customer Wallet Estimation Framework for Sales Territory Optimization

- Trained Quantile Regression GBM models to estimate potential spend bands, enabling tiered wallet segmentation from small-company to enterprise; achieving a 121% lift in revenue correlation over traditional approach based on company's sales volume
- Developed SparkR-based data pipelines for feature engineering, processing 12-month Akamai revenue, network traffic, and D&B firmographics to support model training using Spark MLlib
- Built a real-time scoring pipeline using Salesforce APIs to score accounts at creation, enabling instant sales and marketing activation for account prioritization, share-of-wallet estimation, territory design, and upsell targeting
- Deployed the wallet scoring engine as a systemd daemon on an on-prem server, meeting SOX compliance requirements, as well as created a PL/SQL package for in-database scoring of 300M non-customer companies to support paid search targetting

Education

University at Buffalo, The State University of New York **2015**
Master of Science, Industrial Engineering (with specialization in Operations Research)
Alpha Pi Mu, Industrial Engineering Honor Society, Omega Rho Honors Society GPA: 3.95/4.00

Publications

- A. Kumar et al. "Inferring origin-destination pairs and utility-based travel preferences of shared mobility system users in a multi-modal environment," Transportation Research Part B, 2016, 91(C):270-291
- A. Kumar et al. "Life Cycle Cost Analysis of Ready Mix Concrete Plant," Journal of The Institution of Engineers Series A, 2014, 94(2):229-233

Skills

- **Programming Languages:** Python, R, SQL
- **Tools:** Git, MLflow, Tensorboard, Docker, Ansible, Jenkins CI/CD
- **Databases:** Oracle, PostgreSQL, Neo4j, Memgraph
- **Big Data:** Apache Spark, Hadoop, Ray
- **Cloud Platforms:** Linode