

Prateek Sethi

Senior Machine Learning Engineer

San Francisco, California, U.S.A • prateeksethi814@gmail.com • (540) 824-8990

Senior Machine Learning Engineer with end-to-end ownership of ML models in real-time bidding — from architecture and feature pipelines to production deployment and A/B-validated business impact.

EXPERIENCE

Senior Machine Learning Engineer

RZR (formerly Aarki)

Jan 2025 — Present
San Francisco, California

- Owned the BTI model end-to-end, redesigning it from impression-to-attributed-space modeling to joint attributed + non-attributed (A+NA) space bidding - driving a ~10pp average gain in P80 installer recall in production.
- Designed and implemented multiple Prefect pipelines processing ~200M records/day, feeding both model training data and downstream analytics.
- Identified a persistent mismatch across multiple features between the offline training pipeline (Python) and the online inference feature vectors (Rust); designed and implemented parity detection to catch and fix it, reducing daily mismatch rate from ~50% to <5%.
- Evaluated DCNv2, DeepFM, and DNN offline on 42M+ samples, selected DeepFM as top performer, and validated it via A/B testing before production rollout as the new champion.
- Redesigned iOS user acquisition modeling to separately capture click-driven and no-click (view-through) install intent, replacing a single install model with dedicated CTR + CVR + VTA DNN models; validated gains via A/B tests against the production champion.
- Developed advertiser-specific models for strategic iOS UA accounts, including a filter model that reduced CPI from \$200+ to ~\$60 and a CPA model achieving 0.979 ROC AUC.
- Cut Aerospike join runtimes from ~9 hours to ~1 hour by parallelizing ClickHouse queries and optimizing query design.
- Built LMDB-based versioned feature datasets to guarantee online-offline lookup consistency - ensuring a production model always reads the feature version it was trained on, even as newer feature sets are rolled out.
- Mentored 5+ incoming ML engineers through knowledge-transfer sessions and model documentation covering data pipelines, training workflows, and production systems.

Data Scientist

Aarki

Jan 2024 — Jan 2025
San Francisco, California

- Drove feature development, evaluation, and optimization for ABLR and DeepFM models across install, purchase, and multi-stage prediction objectives for Aarki's mobile advertising DSP.
- Built a scalable Sentence-Transformer + GloVe clustering pipeline to compress 222K+ apps and 100K+ app bundles into semantic groups using non-attributed user data from Aerospike, improving feature coverage for DeepFM models; automated daily ingestion and metadata enrichment.
- Integrated WURFL device (MSRP/model) and Sensor Tower (app-overlap) features into the DNN pipeline, boosting model performance.
- Developed a funnel event model predicting user purchase likelihood from event sequences, improving conversion and revenue outcomes.
- Designed a bid scaling algorithm for adaptive RTB bid-range management and a risk-scoring variable to rank and allocate budget across competing campaigns, enhancing campaign efficiency.
- Automated daily ETL of 8M+ records from ClickHouse and Aerospike into Hadoop/HDFS for model training and feature generation.

Research Assistant, Software and Machine Learning

Commonwealth Cyber Initiative

Dec 2021 — Jan 2024
Arlington, Virginia

- Proposed and validated a Deep Q-Learning agent for adaptive proof-of-work in private blockchain systems, maintaining fairness below 30% within a 20-second constraint.
- Developed ML models for O-RAN slicing with 97%+ accuracy and productionized them as Dockerized Flask microservices on Kubernetes.
- Deployed a 10-node blockchain-enabled federated learning environment to train distributed models with 98%+ accuracy on MNIST.

Software Developer, Data and Analytics

Orange Business Services

Aug 2019 — Jul 2021

Gurgaon, India

- Delivered 6 Python automation solutions that eliminated 40+ manual hours per month across data extraction, reporting, and internal operations, working in an Agile (JIRA) workflow.
- Built automation workflows, dashboards, and 2 NLP-based chatbots using Python, Selenium, REST APIs, SQL, Power BI, and RASA.

EDUCATION

Master of Science in Computer Engineering, Virginia Tech, U.S.A

Aug 2023

Bachelor of Technology - Electronics and Communication, MAIT, GGSIPU, India

Jun 2019

SKILLS

ML & Modeling: Python, Rust, PyTorch, TensorFlow, Keras, scikit-learn, PySpark, pandas, NumPy, SciPy, statistics, A/B testing, deep learning, reinforcement learning, LLMs, RAG.

Data & Pipelines: SQL, ClickHouse, Hadoop, ETL, Prefect, Databricks, Aerospike, LMDB, Elasticsearch, Kibana, Logstash, Redash.

Infra & Tools: Flask, FastAPI, APIs, Docker, Kubernetes, Kafka, Linux, Git, AWS, Azure, OpenStack, Power BI, Tableau, Excel.

CERTIFICATIONS

Microsoft Certified: Azure AI Fundamentals

Apr 2023

AWS Certified Cloud Practitioner

Sep 2022

PUBLICATIONS

P. Sethi et al., "Fair Consensus in Blockchain with Heterogeneous Miners using Reinforcement Learning Aided Adaptive Proof-of-Work," IEEE CCNC 2024.

J. Mallu et al., "AI/ML Data-driven Control Loop for Managing O-RAN SDR-based RANs," IEEE INFOCOM 2023 (Demo).