Vedang Waradpande

+1 (848) 437-1822 vedang.waradpande@gmail.com vedangw.github.io LinkedIn GitHub

EXPERIENCE

Senior Machine Learning Engineer, ***PayPal*** – San Jose, CA, USA ***Mar 2025 – Present***

- Senior MLE on the foundational AI/ML data platform team and a key contributor to the Graph ML vertical, building large-scale Entity Resolution systems and social graphs that power downstream surfaces (Ads, Search) across PayPal and Venmo.
- **Venmo User Search:** Leading modeling and graph schema design for Venmo's typeahead user search (**90M+** users, billions of user-user edges, sub-**100ms** latency), improving the graph-based retrieval and reranking layer of the broader search pipeline.
  - Trained a LightGBM LambdaRank model on search history, candidate popularity, and transaction features to produce user-user graph edge scores (**90% recall@5**); scaling 2-hop neighborhood retrieval with a Redis edge-score cache [WIP].
- **Customer Graph:** Led the development of the incremental pipeline to link billions of PayPal, Venmo, Xoom, and Hyperwallet accounts, writing core production logic using graph-native Spanner Graph lookup, rule-based matching, and graph clustering.
  - Built a 30-minute PySpark pipeline now deployed in production, replacing a third-party vendor solution at 93% alignment and contributing to 8-figure savings by avoiding a multi-year renewal contract.
- **Product Knowledge Graph:** Led a team of 4 on cross-store unification of **100M+** products, owning architecture, metric design, and core implementation for both the bulk and daily incremental pipelines. Stack: BigQuery, Vertex AI, Dataproc, Spanner.
  - Built an end-to-end retrieval → matching → clustering pipeline using vector search, embedding-based and probabilistic ML matching, and large-scale graph clustering, delivering **88%** precision and **58%** recall (**5x** recall improvement over baseline).

Machine Learning Engineer II, ***TikTok*** – Mountain View, CA, USA ***Aug 2024 – Mar 2025***

- Shipped ML workflows for Creator Monetization integrity, combining Transformer-based image embeddings, LightGBM classifiers, and social graph clustering to detect and action spam creators daily; aligned teams on cluster quality metrics across surfaces.
- Optimized human-in-the-loop moderation by shipping ML-based heuristics to isolate high-confidence spam and dynamically route cases, enabling **27%** more cases labeled per day in production and reducing daily capital loss.
- Improved the core spam-classifier training pipeline in PySpark via distributed hyperparameter optimization, feature selection, and conditional HPO based on data shift; reducing the end-to-end training latency from **3+ hrs to < 30 mins**.

Machine Learning Researcher (RTS III), ***Rutgers University*** – Newark, NJ, USA ***Jun 2023 – Aug 2024***

- Developed an ML-based TB drug-discovery screening pipeline using tree-based (XGBoost, LightGBM, CatBoost, Random Forest) and deep learning (CNN, LSTM, GNNs, Transformers/BERT) models, achieving **>95%** test accuracy and **12** promising drug candidates.

Data Science Intern, ***Faire*** – San Francisco, CA, USA ***May 2022 – Aug 2022***

- Built a real-time XGBoost model for Faire's two-sided marketplace (brands ↔ retailers) identifying no-payment-intent retailers, with cross-order score aggregation enabling underwriting actions; addressed **40%** defaults with an estimated **25%** commission increase.

Data Scientist, ***Razorpay*** – Bengaluru, KA, India ***Jun 2019 – Aug 2021***

- Pioneered key fraud-detection components of Razorpay's risk platform: shipped real-time XGBoost + CNN model for incomplete addresses (**94% AUC**) and XGBoost models for e-commerce fraud (**75% AUC**), owning the end-to-end pipeline with Airflow ingestion.
- Designed and built a payment-method recommender using time-series data (**+4%** checkout retention), validated in Spark (**100M+** samples) and served via a Scala/Akka REST API (latency **< 300ms**).

SKILLS

Expertise: Machine Learning, Deep Learning, Recommender Systems, Search & Retrieval, Ranking, Fraud Detection, A/B Testing, NLP, Computer Vision, Large Language Models (LLMs), Retrieval Augmented Generation (RAG), Data Science,
Languages/Tools: Python, SQL, Scala, Bash • PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, Hugging Face, XGBoost, LightGBM
Backend, Data & ML Infra: Airflow, PySpark, Ray, MLFlow, AWS (S3, EC2, SageMaker), GCP (BigQuery, Spanner, Vertex AI, Dataproc)

EDUCATION

M.S. in Computer Science - **Rutgers University, New Brunswick** ***Sep 2021 – May 2023***

B.E. (Honors) in Computer Science - **Birla Institute of Technology and Science, Pilani** ***Aug 2015 – Jul 2019***

PROJECTS

Retrieval Augmented Generation using LLMs for Question Answering, Fact-Checking, and Entity Linking [Code] [Paper]

Visual Question Answering with Generative Image & Language Models [Code] [Paper]

PUBLICATIONS

- "Identification of Antituberculars with Favorable Potency and Pharmacokinetics through Structure-Based and Ligand-Based Modeling." bioRxiv, 2025-02. (First author) [Link]
- "Predicting Completeness of Unstructured Shipping Addresses Using Ensemble Models", SIGIR eCom '21. (First author) [Link]