Khalid Mehtab Khan

khalidmehtabk@gmail.com | 404-263-7813 | San Francisco, CA | https://github.com/Kahl-d

https://www.linkedin.com/in/khalidm-khan/ | https://khalidmk.vercel.app

EDUCATION

San Francisco State University   Aug 2023 – Aug 2025
Master of Science, Data Science & Artificial Intelligence (GPA: 4.0)
  Coursework: Software Engineering, Statistics and Data Mining, Deep Learning, Algorithm Analysis, Artificial Intelligence, Data Visualization.

SKILLS

Programming & Tools: Python, SQL, PyTorch, LangChain, Flask, FastAPI,

Data Science & ML: Predictive Modeling (Classification, Regression), Feature Engineering, Ensemble Methods (Random Forest, XGBoost), t-SNE, Model Evaluation

AI/ LLMs: Transformers (BERT, T5, Mistral-7B), Named Entity Recognition, Fine-Tuning, Retrieval-Augmented Generation (RAG), Semantic Search, Agentic AI

MLOps & Cloud: AWS, Experiment Design & Monitoring, Model Deployment, Vector Databases

PROFESSIONAL EXPERIENCE

Tacit Alma Lab   Aug 2024 – Present | San Francisco, CA
Research Assistant
- Developed a context-aware classifier using DeBERTa with domain-adaptive MLM and BiLSTM-based attention to capture sentence meaning and context, achieving a 4–5% macro F1 gain over BERT baselines.
- Tackled extreme class imbalance and class overlapping by integrating Focal Loss with dynamically weighted sampling, improving stability on minority class sensitivity.
- Established an evaluation protocol with hyperparameter tuning via Optuna, monitoring training dynamics, and auditing predictions, reducing manual annotation workload by 75%.

Innovaccer   Jan 2021 – Sep 2022 | Noida, India
Healthcare Associate
- Led a cross-functional team in analyzing multi-channel outreach campaigns and healthcare summits (HIMSS, VIVE), generating insights that improved lead targeting and boosted ROI.
- Designed HTML-based outreach materials and Tableau dashboards to support data-driven engagement with U.S. payers and providers, increasing sales-qualified leads by 15%.
- Collaborated with data, marketing, and design teams to align messaging, cut data latency by 30%, and drive targeted outreach to senior healthcare stakeholders.

RELEVANT PROJECTS

Secure Sense (Winner, SF Hacks 2025 Emerging AI Innovation)   Mar 2025 – Aug 2025
- Engineered a privacy-centric PII redaction tool as a browser extension using prompt-conditioned LLaMA models for real-time, on-device inference, eliminating cloud reliance and improving data security.
- Applied transfer learning and staged knowledge distillation to shrink the model by 80%, while maintaining 94% PII detection accuracy on limited local compute.
- Employed a serverless contingency framework using AWS Lambda and S3 to manage complex redaction edge cases, supporting hybrid execution paths with 99% operational uptime.

Context-Aware Data Augmentation Tool for Low Data Scenarios   Aug 2024 – Dec 2024
- Fine-tuned BioBERT with masked language modeling and knowledge graph conditioning to generate clinically meaningful synthetic samples for low-resource public health classes.
- Improved XGBoost and Random Forest accuracy by 3.5–5% on imbalanced datasets, outperforming SMOTE and back-translation in precision and recall.