SULMAN A. KHAN

1047 East 14th St., Apt #2, Brooklyn, NY 11230

(718) 755-6739 • sulman@vt.edu • www.linkedin.com/in/sulman-khan • https://sulmank.github.io

SKILLS

Programming languages: Python, SQL (PostgreSQL, MySQL), JavaScript, HTML/CSS
Cloud Platforms: AWS, GCP
Tools & Frameworks: Git, Docker, Terraform, Kubernetes, Apache Airflow, PyTorch, Flask, LangChain
Data Tools: BigQuery, Looker, Apache Spark, dbt
Domains: A/B Testing, Machine Learning, Deep Learning, Natural Language Processing, Recommender Systems, Generative AI, Prompt Engineering, Retrieval-Augmented Generation

WORK HISTORY

Nuage Software Corporation, NY   **June 2024 - Present**
AI/DS Consultant

- Engineered a scalable ETL pipeline on GCP using Apache Airflow to orchestrate workflows, transformed 10K+ rows of daily data via dbt and Apache Spark, optimized BigQuery storage efficiency, and built Looker dashboards to visualize insights, enabling data-driven customer churn prediction and retention strategies.
- Implemented customer churn prediction models, applying advanced feature engineering to drive actionable insights and reduce churn by 30%.
- Consulted on generative AI solution deployment using retrieval-augmented generation (RAG) and prompt engineering to automate document analysis workflows, achieving a 25% accuracy improvement in data extraction.

Fingercramp, NY   **May 2018 - January 2024**
Data Scientist

- Applied decision-based heuristics on player match statistics to develop a character balancing model, effectively doubling the number of characters utilized from 16 to 32 and improving overall game balance.
- Established and maintained a PostgreSQL database for match statistics, enabling complex querying across multiple tables and schemas for in-depth data analysis and reporting.
- Designed data visualization dashboards for a streaming platform, driving a 48% increase in peak viewership by improving user engagement.

PERSONAL PROJECTS

Reddit AI Pulse: Building an AI-Powered Data Pipeline   **Winter 2024**
https://github.com/SulmanK/reddit_ai_pulse_cloud_public

- Designed and implemented a scalable ETL pipeline on Terraform-provisioned GCP infrastructure, leveraging Apache Airflow for orchestration, GCS as a data lake, dbt and Apache Spark for transformations, and BigQuery for warehousing; automated daily processing of 1K+ AI-focused subreddit comments and ensured reliability with Grafana, Prometheus, MLflow, and Cloud Monitoring.
- Improved content analysis efficiency by 25% by deploying a multi-task NLP pipeline (RoBERTa, BART, Gemini AI) in a React-based dashboard, processing 1k+ daily Reddit discussions to generate dynamic visualizations and actionable summaries that accelerated data-driven strategy decisions.

RecSys Challenge 2024   **Fall 2024**
https://github.com/SulmanK/2024-Recsys-Challenge

- Engineered the EB-NeRD dataset for the Ekstra Bladet RecSys Challenge 2024, processing 10M+ user interactions and 500K+ articles with embeddings, session metadata, and categorical encodings to build a high-quality training corpus for CTR prediction models.
- Ranked Top 100 by optimizing personalized news recommendation models, leveraging feature selection and hyperparameter tuning to improve click-through rate prediction ROC-AUC by 37.5% ($0.500 \rightarrow 0.6875$).

eBay: Phone Auction Aide   **Fall 2020**
https://github.com/SulmanK/eBay-web-crawler-phone-auctions

- Engineered a Python web scraper to automate eBay phone auction data collection, using AWS Lambda to trigger Batch jobs for scalable data processing, orchestrated via EventBridge scheduling, and storing results in PostgreSQL—reducing manual effort by 40% while maintaining 99% uptime.
- Built a real-time auction analytics dashboard with Dash to visualize price trends, bid velocity, and auction competitiveness, enabling users to identify high-value opportunities 30% faster and optimize bidding strategies.

EDUCATION

Stony Brook University, Stony Brook, NY   **May 2018**
Master of Science, **Electrical Engineering (Concentration in Machine Learning Systems)**

Virginia Polytechnic Institute and State University, Blacksburg, VA   **May 2016**
Bachelor of Science, **Materials Science and Engineering**