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
Tools & Frameworks: Git, Docker, Kubernetes, AWS, GCP, Apache Airflow, BigQuery, Looker, Apache Spark, dbt, PyTorch, Flask, LangChain
Domains: 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

- Designed and integrated an ETL data pipeline on GCP, using Apache Airflow for orchestration, PostgreSQL for the database, dbt for data transformation, BigQuery for data warehousing, and Looker for visualization, to collect and process user data, enabling customer churn prediction and supporting data-driven retention strategies.
- Implemented a machine learning model for customer churn prediction, focusing on advanced feature engineering and leveraging predictive insights to enhance player retention strategies by reducing churn.
- Advised clients on implementing generative AI methodologies, including retrieval-augmented generation, prompt engineering, and AI agentic workflows, to develop customized AI solutions tailored to specific business challenges.

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.
- Produced data visualization dashboards for the streaming platform, resulting in a 48% increase in max concurrent viewership by enhancing the user experience.

Personal Projects

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

- Engineered and preprocessed the large-scale EB-NeRD dataset for the Ekstra Bladet RecSys Challenge 2024, integrating user interaction logs, session metadata, and enriched article features to prepare high-quality data for machine learning models.
- Applied and optimized personalized news recommendation models for click-through rate prediction, leveraging feature selection, recommendation techniques, and ROC-AUC evaluation. Achieved a top 100 placement in the Ekstra Bladet RecSys Challenge 2024.

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

- Created a Python-based web scraper to gather phone auction data from eBay, implementing workflow services for automation and maintenance of a PostgreSQL database, which resulted in a 25% improvement in efficiency and throughput.
- Launched a user-friendly application for real-time monitoring of phone auctions, providing valuable metrics to assist in auction selection, and improving the user experience.

Video Game Recommendation Engine **Summer 2020**
https://github.com/SulmanK/Video-Game-Recommendation-Engine

- Developed a Python-based parsing tool to aggregate video game data from the Giant Bomb API and efficiently store entries into a PostgreSQL database, ensuring streamlined access and management of large datasets.
- Utilized NLP algorithms including TF-IDF vectorization and cosine similarity to develop a content-based recommendation system for suggesting relevant video games.
- Deployed an application that allows users to input video game titles and receive personalized recommendations, enhancing user engagement.

Education

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

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