KUSHAGRA KAUSHAL

Github linkedin

SUMMARY

Accomplished Data Scientist and Engineer with over three years of professional experience in data engineering, advanced analytics, machine learning, and applied research. Demonstrated expertise in designing and implementing comprehensive data solutions, conducting sophisticated data analyses, and developing high-impact predictive models. Committed to harnessing data-driven insights to optimize business performance and deliver strategic value.

EDUCATION

Indian Institute of Information Technology Bhopal (India)
B.Tech Information Technology

SKILLS

- **Programming Languages:** Python, C++, C
- **Machine Learning Frameworks:** Scikit-learn, TensorFlow, PyTorch, Deep Learning, Computer Vision, Natural Language Processing
- **Data Engineering Technologies:** dbt, Apache Airflow, Google BigQuery, Cloud SQL
- **Cloud Platforms:** Google Cloud Platform, Amazon Web Services
- **Data Visualization Reporting Tools:** Looker Studio, Plotly, Dash
- **DevOps MLOps Technologies:** Kubernetes, Docker, Devtron CI/CD
- **Additional Technologies:** Linux, MySQL, Bash Scripting, FastAPI, Flask, Django

WORK EXPERIENCE

Data Scientist New Engen - Seattle, US (Remote) November 2022 - Present
`https://www.newengen.com/`

- Architected and deployed advanced AI-powered chatbot systems utilizing multiple large language models (Claude, Gemini) to perform sophisticated reasoning and deliver actionable marketing insights through function calling and optimized content caching.
- Designed and maintained robust data pipelines leveraging Adverity, custom Python scripts, dbt, BigQuery, and Cloud SQL on Kubernetes, implementing ETL workflows that enhanced real-time analytics ef ciency by approximately 60%.
- Conducted comprehensive data analysis on 5 terabytes of marketing data in BigQuery, deriving critical KPIs including spend, revenue, ROAS, CPC, and AOV to inform strategic decision-making.
- Developed dynamic and scalable reporting dashboards in Looker Studio, integrating data from diverse sources (GA4, Meta, Google, etc.) to support data-driven decisions, contributing to the LIFT SaaS platform.
- Engineered machine learning solutions, including a pacing and budgeting forecast system and a recommendation engine, utilizing real-time data analysis to enhance operational ef ciency.

Data Science Consultant Kauriink Pvt. Ltd. - New Delhi, India (Remote) August 2022 - November 2022
`https://www.techatplay.ai/`

- Developed and validated deep learning models for automated player performance analysis using computer vision, achieving 87% accuracy in posture classi cation (6 classes) and 92% accuracy in shot type classi cation (8 classes).
- Implemented advanced color segmentation and sliding window techniques to optimize object tracking in video data.
- Collaborated on the integration of transfer learning models (Mediapipe, YOLO) to accurately detect player posture, movement, ball trajectory, and shot type.
- Exhibited exceptional problem-solving and analytical capabilities, applying technical expertise and innovative approaches to deliver high-impact solutions.

RESEARCH EXPERIENCE

Research Assistant University of Michigan - Michigan, US (Remote) May 2023 - Dec 2023

- Designed an encoder-decoder neural network architecture to generate high- delity simulated LiDAR point cloud data from 3D scenes.
- Leveraged Blender and Python automation for ef cient dataset creation, incorporating advanced object translation and scaling techniques.
- Engineered a comprehensive model pipeline, encompassing data pre-processing, PointNet-based encoder, decoder, and a tailored loss function utilizing a modi ed Chamfer Distance metric.

Research Intern IIT Indore - Indore, India (Remote) May 2022 - February 2023

- Developed advanced machine learning models to predict speaker attributes (age, height, sex) from audio recordings with high accuracy.
- Extracted and normalized acoustic features (MFCC, Mel- lterbank) using CMVN techniques to enhance model performance.
- Applied sophisticated audio data augmentation methods, including speed perturbation and masking, to improve model robustness.
- Designed and trained models leveraging transformers, Wav2Vec, and SincNet, achieving performance comparable to or surpassing state-of-the-art benchmarks.

PROJECTS

Obelisk Ongoing

`https://github.com/kshgrk/obelisk`

- Developed Obelisk, a real-time chat application leveraging Temporal workflows for fault-tolerant orchestration and OpenRouter's free AI models (e.g., DeepSeek, Mistral, Llama), enabling dynamic model switching mid-conversation without context loss.
- Architected a multi-layered system using FastAPI for proxy and backend services, Vanilla JavaScript frontend, and SQLite database, facilitating seamless session persistence, real-time streaming responses, and live updates.
- Implemented an extensible tool registry for dynamic tool calling, including built-in integrations for calculators and weather APIs, with a plugin-based architecture allowing easy addition of custom tools at runtime.
- Ensured robustness through Temporal's retry mechanisms and layered design, supporting resumable conversations, model-agnostic sessions, and integration with external services for enhanced functionality.

IntelliCodebase Jan 2025

`https://github.com/kshgrk/IntelliCodebase.git`

- Architected an advanced LLM-powered system (utilizing Gemini) for comprehensive codebase analysis and modi cation.
- Integrated function calling with Bash and Python to execute system-level tasks, including le creation, deletion, and script execution.
- Implemented an optimized content caching mechanism to enhance performance during analysis of large-scale codebases.
- Developed functionality to identify codebase issues and propose targeted xes, with selective le analysis capabilities.

LSMTree-AVL Nov 2024

`https://github.com/kshgrk/LSMTree AVL.git`

- Engineered a Python-based LSM-tree database utilizing AVL trees for ef cient memtable and index management, optimizing storage and retrieval operations.
- Incorporated a Bloom lter for rapid key presence veri cation, memtable flushing to SSTables, and multi-threaded compaction strategies.
- Developed a write-ahead log mechanism to ensure data durability and robust recovery processes.