Kushagra Kaushal

India · kushagra9nov@gmail.com · +91 9599958093 · Github · linkedin

Summary

Data Scientist and Engineer with over **2.5 years of experience** in diverse areas including data engineering, analysis, reporting, machine learning, and research. Proven ability to build and deploy end-to-end data solutions, analyze large datasets, and develop impactful models. Passionate about leveraging data to drive insights and improve business outcomes.

Education

Indian Institute of Information Technology Bhopal
B.Tech Information Technology

Skills

- **Programming Languages:** Python, C++, C
- **Machine Learning:** Scikit-learn, TensorFlow, PyTorch, Deep Learning, Computer Vision, NLP
- **Data Engineering:** dbt, Apache Airflow, BigQuery, Cloud SQL
- **Cloud Computing:** GCP, AWS
- **Data Visualization & Reporting:** Looker Studio, Plotly, Dash
- **DevOps & MLOps:** Kubernetes, Docker, Devtron CI/CD
- **Other Tools & Technologies:** Linux, MySQL, Bash Scripting, FastAPI, Flask, Django

Work Experience

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

- Experience in building complex AI chatbots using multiple LLMs (ChatGPT, Gemini) to perform complex reasoning and analysis on marketing data to provide actionable insights by doing function calling and content caching.
- Engineered and maintained data pipelines using Adverity, custom Python scripts, dbt, BigQuery, Cloud SQL on Kubernetes; including ETL workflows for real-time analytics.
- Performed data analysis on terabytes of marketing data in BigQuery, calculating KPIs like spend, revenue, ROAS, CPC, and AOV.
- Developed dynamic, scalable Looker Studio reports from various sources (GA4, Meta, Google, etc.) for data-driven decisions, contributing to LIFT SaaS product with reporting dashboards and data integration.
- Developed ML tools including a pacing and budgeting forecast system and a recommendation engine using real-time data analysis.

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

- Developed and evaluated deep learning models for automated player performance analysis using computer vision techniques on video data, achieving 87% accuracy for posture classification (6 classes) and 92% accuracy for shot type classification (8 classes).
- Implemented color segmentation and sliding window approaches for efficient object tracking in video footage.
- Collaborated on integrating transfer learning models (Mediapipe, YOLO) to detect player posture, movement, ball movement, and shot type.
- Demonstrated strong problem-solving and analytical skills by applying technical expertise, critical thinking, and creative solutions to deliver impactful results.

Research Experience

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

- Developed an encoder-decoder neural network to generate simulated LiDAR point cloud data from 3D scenes.
- Utilized Blender and Python automation for dataset generation, incorporating object translation and scaling.
- Designed the complete model pipeline, including pre-processing, encoder (PointNet), decoder, and a custom loss function (modified Chamfer Distance).

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

- Developed models to predict speaker age, height, and sex from audio recordings.
- Extracted acoustic features (MFCC, Mel-filterbank) and applied normalization techniques (CMVN).
- Implemented audio data augmentation techniques (speed perturbation, masking).
- Designed and trained models using transformers, Wav2Vec, and SincNet, achieving results comparable to or exceeding state-of-the-art in some cases.

Projects

IntelliCodebase Ongoing

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

- Developed an LLM-powered system (using Gemini) for codebase analysis and modification.
- Implemented function calling with Bash and Python for system-level tasks (file creation/deletion, script execution).
- Integrated a content caching mechanism to optimize performance during analysis on large codebases.
- Analyzes codebase for issues and suggests fixes, with an option to analyze specific files.

LSMTree-AVL Nov 2024

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

- Implemented a Python-based LSM-tree database utilizing AVL trees for the memtable and index for efficient storage and retrieval.
- Included a Bloom filter for quick key presence checks, memtable flushing to SSTables, and compaction strategies on a separate thread.
- Developed a write-ahead log for data durability and recovery.

Tensorflow Object Detection June 2022

`https://github.com/kshgrk/tf-object-detection`

- Implemented real-time object detection using TensorFlow 2 with SSD MobileNet v2.
- Generated a real-time dataset using a webcam and created image annotations with LabelImg.