Karush Pradhan

Data Scientist

Data Scientist with hands-on experience in Python and SQL, specializing in time-series analysis, ML modeling (XGBoost), and dashboard development with Streamlit. Skilled in building end-to-end data pipelines, extracting and transforming raw datasets, and deploying predictive models using Google Cloud Functions. Passionate about applying data science to real-world systems like HVAC control, energy analytics, and automation. Background in accounting supports sharp business insight and KPI-focused analysis.

Experience

Data Scientist — **Menteru Inc · Tokyo, Japan**, Oct 24 - present

Architected and maintained modular, production-grade data pipelines using Python for large-scale time-series sensor data. Leveraged Cloud Storage for scalable data ingestion, and orchestrated stream and batch processing with Cloud Dataflow and Pub/Sub for message-driven workflows. Designed outlier detection, interpolation, feature engineering, resampling, and data transformations using Pandas and NumPy.

Extracted and transformed raw sensor and metadata from relational databases using Cloud SQL (MySQL/PostgreSQL) and BigQuery. Performed complex joins and aggregations to support downstream model-ready datasets.

Trained, tuned, and evaluated XGBoost models for HVAC performance simulation in Vertex AI, optimizing for energy efficiency, thermal comfort, and productivity. Included structured feature selection, EDA, and model evaluation (e.g., RMSE, residual analysis) using managed model training and evaluation pipelines.

Developed and deployed interactive multi-tab dashboards using Looker Studio or embedded Looker alongside Streamlit for dynamic visualization of KPI trends, control logic effectiveness, and environmental metrics—enabling impactful data storytelling for stakeholders.

Automated real-time device actuation via REST API integrations using Google Cloud Functions. Connected predictive logic with building systems (SwitchBot) for dynamic, serverless control.

Collaborated closely with cross-functional teams, combining cloud infrastructure (Compute Engine, Cloud Functions), data engineering (Dataflow, BigQuery), and business stakeholders to ensure alignment between technical solutions and operational objectives.

Previous Experience – Accounting & Finance (2014–2024)

Various Roles | Sydney, Australia & Kathmandu, Nepal
Conducted financial analysis, budgeting, and reporting across healthcare, aesthetics, and public sector organizations.
Reconciled and extracted large volumes of transactional data to identify trends, anomalies, and process improvements.
Managed collections, credit evaluation, aged debt, and bad debt forecasting across multi-client environments.
Led process audits, mentored junior staff, and contributed to internal tool and system development for AR/AP efficiency.

Organizations:

LifeHealthcare	Accounts Receivable Coordinator	*Sydney, Australia*	Jul 2023 – May 2024
Allergan Aesthetics	Accounts Receivable Coordinator	*Sydney, Australia*	Jul 2022 – Jul 2023
Medlab Pathology	Accounts Receivable	*Sydney, Australia*	Jun 2021 – Jul 2022
Mastercare Australasia	National Credit Manager	*Sydney, Australia*	Mar 2019 – Mar 2021
Bajra Technologies Pvt. Ltd.	Account Manager	*Kathmandu, Nepal*	Jul 2014 – Oct 2015

Selected Data Science Projects

Resume Chatbot — AI & Full Stack Application
github.com/karushp/resume-chatbot https://karushp.github.io/resume-chatbot/
An AI-powered chatbot that answers questions about my resume using RAG (Retrieval-Augmented Generation)
Tools: Python, Flask, FAISS, Cohere API, Groq API, HTML/CSS/JavaScript, Github Pages

EMS Energy Consumption Dashboard — Energy Management System Dashboard
github.com/karushp/ems-dashboard
An interactive Streamlit dashboard for analyzing energy consumption data across Kansai and Kanto regions in Japan, featuring dynamic filtering and comprehensive visualizations.
Tools: Python, Streamlit, Pandas, Plotly, Parquet

YEN-USD Analysis and Prediction Model — Data analysis and Prediction model
github.com/karushp/yen-usd-analysis
Exploratory analysis of YEN to USD and prediction using Random Forest Classifer
Tools: Python, Scikit-learn, Pandas, Seaborn, Jupyter

Heart Disease Prediction — Data analysis and Prediction model
github.com/karushp/heart-disease-predict
This project predicts heart disease, optimizing K-Nearest Neighbors through scaling, normalization, and grid search.
Tools: Python, Scikit-learn, Jupyter Notebook, Pandas, Seaborn

Safety Helmet Detection — Computer Vision Model
Real-time detection ML Model identifying instances of safety helmet usage.
Tools: YOLO, OpeanCV, Python, Streamlit

Contact

karushpradhan@gmail.com
Mabashi, Chiba 271-0046

Socials

linkedin.com/in/karushpradhan
github.com/karushp

Education

Le Wagon Tokyo
Data Science and AI Bootcamp
2024

Macquarie University, Syndey (Australia)
Master in Accounting (CPA Extenstion)
2018

Christ University, Bangalore (India)
Bachelor of Business Management specialized in Marketing
2014

Skills

Programming & Tools: Python, SQL, Pandas, NumPy, Jupyter, Git, GCP (Cloud Functions)

Machine Learning: Scikit-learn, XGBoost, KNN, Random Forest, Time-series forecasting, Hyperparameter tuning, Cross-validation, Model evaluation (RMSE, R^2, residuals), Wellness-based control logic

Visualization: Streamlit, Matplotlib, Seaborn, Looker Studio, Plotly, Custom KPI dashboards, Multi-tab UI design

Cloud & Automation: Google Cloud Functions, Pub/Sub, Cloud Dataflow, BigQuery, SwitchBot API, REST APIs, GCP IAM & roles

Data Engineering & ETL: Data preprocessing, Feature engineering, Outlier detection, Interpolation, Resampling, Time alignment, Multi-sensor data handling, Class-based pipeline design

Languages

English (Fluent)
Nepali (Native)

Additional

Holds valid Japanese Driver's License (普通自動車免許)