

# Devansh Pathak

## Health Data Scientist

Open to Any Location in US | devansh@pathak.dev | (913) 256-8636 | linkedin.com/in/dp817

### SUMMARY

Results-oriented Data Scientist with over 3 years of experience transforming datasets into tangible business value. Expertise in statistical modeling, machine learning, and architecting scalable cloud solutions to automate data processing and analysis. Specialized in building predictive models and deploying them as interactive tools for stakeholders to increase operational efficiency and revenue.

### EDUCATION

#### Master in Health Data Science and Informatics (MS)

University of Kansas Medical Center • Kansas City, Kansas

#### Bachelor of Medicine, Bachelor of Surgery (MBBS)

Smt. NHL Municipal Medical College • Gujarat, India

### SKILLS

**Languages:** Python, R, SQL, SAS

**Machine Learning:** Supervised (Logistic Regression, XGBoost, RF), Unsupervised (Autoencoders, Clustering), Deep Learning (Keras, TensorFlow)

**Data Infrastructure:** ELT/ETL, AWS (S3, Lambda, RDS), SQL Optimization, Data Validation Pipelines, Microsoft Azure, REDCap

**Visualization Platforms:** Tableau, Power BI, JMP Pro

**Tools:** Git, Jupyter, VSCode, REST APIs

### EXPERIENCE

#### Data Science Researcher

Stellatus Services

Sep 2024 – Sep 2025

- Led the end-to-end development of a patient readmission risk model, from data cleaning and feature engineering of a large healthcare dataset to final model evaluation and deployment.
- Improved the identification of at-risk patients by 100% (recall increased from 0.10 to 0.20) by implementing an XGBoost classifier over a baseline Random Forest model.
- Presented a comprehensive business case for the model, demonstrating its potential to improve patient care and generate substantial savings by preventing costly, unplanned readmissions.

#### Data Science Internship

Stellatus Services

Jul 2024 – Sep 2024

- Architected a novel, unsupervised deep learning fraud detection system using an autoencoder, improving the AUC-ROC score to 0.902 from baseline's 0.875, demonstrating superior ability to distinguish between fraudulent and legitimate transactions.
- Engineered the solution to achieve a critical business target of 80% recall, ensuring capture of fraudulent activity to minimize financial loss, a primary objective for risk management in financial services.
- Increased model precision by 38% over the baseline at the target 80% recall (25.4% vs 17.7%), significantly reducing false positives and protecting legitimate customers from disruptive, unnecessary transaction declines.

#### Clinical Data Assistant

Westfield Urgent Care

Jun 2021 – Dec 2022

- Improved patient wait time by 40% using a time-series predictive model that identified peak congestion patterns, leading to redesign of triage workflow, re-allocation of clinic resources, increasing daily patient capacity and nearly doubling clinic revenue.
- Reduced data entry errors by 25% from standardizing physician charting templates via centralized data pipeline and automated data validation checks to clean intake data for 100+ daily patients.

---

## PROJECTS

---

### Patient Risk Prediction & Clinical Dashboard

- Architected a serverless pipeline on AWS (S3, Lambda, RDS) to process a 6,000-patient dataset, engineering features from patients baseline (smoking history) and automated the identification of data errors to reduce time-to-analysis.
- Implemented a Random Forest Classifier (scikit-learn) to proactively identify patients at high-risk; validated model's 68% accuracy by confirming features aligned with clinical risk factors making them both accurate and interpretable.
- Deployed a live and interactive dashboard (streamlit), translating complex outputs into actionable tool, enabling clinical teams to implement targeted monitoring strategies and potentially improve patient safety outcomes.
- **Tech Stack:** Python, SQL, AWS (S3, Lambda, RDS), scikit-learn, pandas, NumPy, Streamlit

### Geospatial Health Determinants Modeling

- Designed a nationwide spatial analysis pipeline in R to model public health outcomes, integrated data from CDC and Census Bureau APIs; corrected geographic clustering with spatial error model (spdep, spatialreg) of 3,068 counties.
- Developed an inferential model (R-squared = 0.67) that identified county-level poverty, creating a framework to identify and stratify geographic hotspots for targeted resource allocation and population health interventions.
- Validated the model (VIFs via car, Cook's distance) and validated coefficient stability across 4 sensitivity analyses, delivering the project in a fully reproducible environment with renv.
- **Tech Stack:** R, tidyverse, spdep, spatialreg, car, sf, ggplot2, Census API, CDC API, renv

### 30-Day Readmission Prediction

- Built and evaluated multiple ML models (Logistic Regression, k-NN, Random Forest, XGBoost) to predict 30-day hospital readmissions, achieving an AUC of 0.79 with the final XGBoost model.
- Engineered features from a 100,000-record dataset and addressed class imbalance, boosting recall for the positive class from 0.10 (Random Forest) to 0.20 (XGBoost).
- Translated complex analytical findings into actionable business insights, presenting the results to stakeholders and showcasing the potential for significant cost savings.
- **Tech Stack:** Python, SQL, scikit-learn, XGBoost, pandas, NumPy, matplotlib, imbalanced-learn

---

## CERTIFICATIONS

---

**NCA Generative AI LLM** • NVIDIA (2025)

**ML Specialization** • Stanford (2025)

**Data/Specimens Only Research** • CITI (2025)

**GCP for Clinical Trials** • CITI (2024)

**Data & Safety Monitoring** • CITI (2024)

**IBM Professional Data Science** • IBM (2024)

---

## PUBLICATIONS & RESEARCH

---

### Mahatma Gandhi Cancer Hospital

- Desai A, **Pathak D**, Pol J, Kadkol GA, Sabnis AL. De-Novo Malignant Peripheral Nerve Sheath Tumor of Adrenal Gland with Divergent Differentiation: Case Report & Literature Review. (No DOI available)

### University of Kansas Medical Center

- Chakravarthi VP, Ghosh S, Dai E, **Pathak D**, Rumi MAK. Transcriptome datasets of ESR2-regulated genes in rat granulosa cells during gonadotropin- induced follicle maturation. Data Brief. 2020;30:105405. doi:10.1016/j.dib.2020.105405
- Chakravarthi VP, Ghosh S, Roy R, Dai E, **Pathak D**, Rumi MAK. Transcriptome datasets of gonadotropin-induced ESR2-regulated genes in rat oocytes. Data Brief. 2019;27:104786. doi:10.1016/j.dib.2019.104786
- Khristi V, Ratri A, Ghosh S, **Pathak D**, et al. Disruption of ESR1 alters expression of genes regulating hepatic lipid & carbohydrate metabolism in male rats. Mol Cell Endocrinol. 2019;490:47–56. doi:10.1016/j.mce.2019.04.005