Devansh Pathak
Data Scientist – Machine Learning, Statistical Modeling, and Data Management
Overland Park, Kansas dpathak.md@gmail.com (913) 256 8636 linkedin.com/in/dp817 github.com/ml-dp

EDUCATION

Master of Healthcare Data Science (MS)
University of Kansas Medical Center • Kansas City, Kansas • Jul 2025
Bachelor of Medicine, Bachelor of Surgery (MBBS)
Smt. NHL Municipal Medical College • Gujarat, India • Aug 2017

EXPERIENCE

Clinical Data Assistant
Westfield Urgent Care **Jun 2021 – May 2023**
- Improved patient wait time by 40% using time-series predictive model that identified peak congestion patterns, leading to redesign of triage workflow and increasing daily patient capacity.
- Reduced data entry errors by 25% from standardizing physician charting templates by developing an ELT pipeline and automated data validation checks to clean intake data for 100+ daily patients.
- Led a data-first resource re-allocation by analyzing clinic revenue and operational data; presented findings to stakeholders which optimized staffing schedule and improved daily clinic revenue by nearly 2x.

USMLE/COMLEX Instructor
Institute of Medical Boards **Oct 2020 – Dec 2021**
- Built performance dashboards and automated reporting that featured heatmaps and trend forecasting, enabling real-time feedback loops leading to an average score improvement of 80 points across 8 cohorts.
- Implemented a predictive analytics framework in Excel and Python to analyze student performance data; identified key predictors of exam success to create personalized learning pathways for over 200 students.
- Utilized a A/B testing framework to evaluate efficacy of different teaching methodologies, resulting in adoption of new curriculum that increased student pass rates by 12%.

Volunteer Research Assistant
University of Kansas Medical Center **Mar 2019 – Sept 2020**
- Maintained a relational database (PostgreSQL) for high-dimensional genomic data, using advanced queries to retrieve and join count matrices with phenotypic data for analysis.
- Performed statistical analysis of RNA-sequencing data in R, utilizing DESeq2 to identify differentially expressed genes and performing pathway enrichment analysis that led to 3 peer-reviewed publications.

PROJECTS

Predictive Risk Modeling for Patient Safety
- Architected a serverless pipeline on AWS (S3, Lambda, RDS) to process a 6,000-patient dataset, engineering features from patient's baseline (smoking history) and automated the identification of data errors to reduce time-to-analysis.
- Developed 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.

Ecological Analysis of Socioeconomic Health Determinants
- Engineered 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 model through comprehensive diagnostics (VIFs via car, Cook's distance) and validated coefficient stability across 4 sensitivity analysis, delivering the project in a fully reproducible environment with renv.

Clinical Risk Stratification from Stress Test Data
- Developed a predictive framework to stratify patient risk for myocardial ischemia, leveraging a multiple linear regression model that explained 40.5% of the outcome's variance (Adjusted R-squared) in 303 patient records.
- Improved model validity by diagnosing significant heteroscedasticity (Breush-Pagan test) and correcting its effects with standard errors (HC3) yielding valid and reliable coefficients for clinical predictors.
- Translated model insights into a clinical decision-making framework, demonstrating that ST segment slope ($p<0.001$) and age ($p<0.013$) were critical predictors, providing physicians to refine pre-test risk assessments.

CERTIFICATIONS

NCA Generative AI LLM • NVIDIA (2025)
Machine Learning Specialization • Stanford (2025)
Data or Specimens Only Research • CITI (2025)
Biomedical Research with Children • CITI (2025)
IBM Professional Data Science • IBM (2024)
GCP for Clinical Trials (US FDA Focus) • CITI (2024)
CRC Foundations • CITI (2024)
Biomedical PI• CITI (2024)
Data & Safety Monitoring in Human Subjects • CITI (2024)
Biomedical Researchers • CITI (2024)