Paritosh Kiran Gandre | Data Scientist

Kent, OH | +1 330 554 8445 | paritoshkrcg@gmail.com | Portfolio | Github | Linkedin

Education

M.S., Data Science, Kent State University, OH — Aug 2023 – May 2025
Relevant coursework: Artificial Intelligence (TensorFlow, PyTorch, Keras), Statistical Learning, Applied & Computational Statistics
B.E., Computer Engineering, University of Mumbai — Jul 2020 – Jul 2023

Technical Skills

- **Languages:** Python, R, SQL
- **Machine Learning & AI:** scikit-learn, TensorFlow, PyTorch, LLMs, Prompt Engineering, Cross-Validation, Model Interpretability
- **Data Engineering & Cloud:** AWS (S3, EC2, Lambda, IAM), Airflow, Docker, Terraform (basic), CI/CD (GitHub Actions), Data Modeling, Governance, Lifecycle Management
- **Visualization & Analytics:** Power BI, Tableau, matplotlib, seaborn, Excel (Power Query, Pivot Tables, VBA)
- **Healthcare Data & Compliance:** FHIR, HL7 (basic), HIPAA awareness, Clinical Data Integration
- **Other:** Git, Agile/Scrum, Unit Testing, REST APIs (Django/Flask)

Experience

Research Assistant – Deep Learning for Genomic Data Analysis — Feb 2025 – May 2025
Kent State University, Kent, OH

- Engineered scalable ETL pipelines in Python and Docker to process genomic and clinical-like datasets, reducing manual data handling by 80%
- Achieved 97% classification accuracy on 40 K DNA sequences using CNN models (TensorFlow/PyTorch) validated via ROC curves and confusion matrices
- Applied Bayesian modeling for interpretable motif discovery and aligned findings with biological significance in collaboration with research faculty

Data Scientist – AI Engineering Intern — May 2024 – Dec 2024
Inke (SafeSpace), Orlando, FL

- Deployed AI models to AWS EC2 via Airflow-driven CI/CD pipelines, improving deployment reliability by 30% and standardizing model retraining
- Built recommendation and sentiment analysis systems (spaCy, NLTK) achieving 85% accuracy and enhancing user engagement for 100 K + requests daily
- Optimized Django REST APIs using Docker and AWS networking best practices, reducing latency by 30%
- Collaborated in Agile sprints and code reviews to ensure quality, maintainability, and unit test coverage

Data Analyst – Supply Chain — Jan 2025 – May 2025
Kent State University, Kent, OH

- Forecasted ingredient demand using ARIMA models on 3 years of POS data, reducing over-purchasing by 15% ($10 K savings)
- Automated SQL extractions and AWS S3 integration for Power BI dashboards, cutting report latency by 30%
- Applied A/B testing for procurement strategies, improving forecast reliability and cost variance control

Operations Data Analyst (Catering Analytics Associate) — Apr 2024 – Dec 2024
Kent State University Dining Services, Kent, OH

- Analyzed 15 K + POS and inventory records using pandas and Hadoop to optimize procurement, reducing food costs by 12% ($4,800 savings)
- Developed Power BI/Tableau dashboards for executive KPIs, accelerating decision-making by 30%
- Created Excel/VBA automations reducing report preparation time by 25%

Projects

Machine Learning & Artificial Intelligence

- Pose Corrector App (Python) / Pose-Corrector-Expo-App (React Native + Mediapipe) — *May 2025 – Present*
  - Developing a cross-platform computer vision app using Mediapipe Pose Estimation and React Native camera integration for real-time form correction
  - Implemented joint angle calculations and feedback logic; optimized TensorFlow Lite inference achieving <50ms latency on mobile devices
- Music Genre Classifier — *May 2025*
  - Built CNN classifier using librosa for MFCC and spectrogram feature extraction; applied batch normalization and dropout
  - Achieved 92% accuracy using stratified k-fold cross-validation and grid search for hyperparameter optimization

- AI-Powered Blog Post Generator with Daily Automation *June 2025*
  - Automated blog generation pipeline using OpenAI GPT API and cron-scheduled Python scripts for topic generation and posting
  - Integrated TF–IDF and cosine similarity for NLP-based topic clustering and relevance scoring
- AI–Image Classification using CNNs (Flood vs Earthquake) *Aug 2024 – Dec 2024*
  - Built CNN model on 2.4K+ labeled images using TensorFlow/Keras; achieved 90.19% validation and 89.9% test accuracy
  - Applied data augmentation, dropout, and learning-rate tuning; validated results via confusion matrix and ROC–AUC
- Genomic Sequence Classification (CNN–HMM Hybrid) *Feb 2025 – May 2025*
  - Designed hybrid CNN–HMM model to classify synthetic DNA motifs into 4 classes with 97% accuracy
  - Compared log-likelihood scoring from HMMs vs convolutional feature maps; validated via cross-entropy loss and F1 metrics
- Image Spam Detection *July 2021 – Apr 2022*
  - Trained CNN and logistic regression models for spam image classification; achieved >95% precision and recall
  - Engineered pixel-intensity and histogram-based features; evaluated performance using ROC–AUC and confusion matrices

Bioinformatics & Healthcare Data Science

- iPSC Differentiation Predictor *Oct 2025*
  - Trained Gradient Boosting and Random Forest models on gene expression data to predict stem-cell differentiation outcomes
  - Used SHAP for feature interpretability and PCA for dimensionality reduction; achieved >85% balanced accuracy
- Healthcare Monitoring System with Comprehensive Dashboard *Jan 2025 – May 2025*
  - Built FastAPI + Dash web app to predict heart and diabetes risk using Random Forest, XGBoost, and CNNs on synthetic data
  - Integrated SQLite backend and real-time visual dashboards; achieved $R^2=0.89$ for heart-disease risk prediction
- Bioinformatics Meets Programming (UCSD Course) *Aug 2025*
  - Implemented core bioinformatics algorithms (Greedy Motif Search, Gibbs Sampling, k-mer counting) in Python
  - Analyzed motif patterns and GC-skew across genomic sequences; validated replication origins through statistical motif scoring

Data Science & Predictive Analytics

- Stock Market Big Data Visualization Project *Jan 2024 – May 2024*
  - Built Python dashboard for stock volatility forecasting using Monte Carlo simulation and Random Forest regression
  - Used Bootstrap resampling for uncertainty quantification; visualized returns distribution and Sharpe ratio variation
- Car Price Prediction (Statistical Learning) *Jan 2024 – May 2024*
  - Applied Linear Regression, Random Forest, and Gradient Boosting on 8K+ listings; achieved $R^2=0.74$
  - Engineered features with one-hot encoding and variance inflation checks; validated via residual analysis and F-test
- Computational Statistics – Airbnb NY 2019 Dataset *Aug 2023 – Dec 2023*
  - Modeled Airbnb prices via linear and quadratic regression using R; applied bootstrapping for confidence intervals
  - Compared borough-level models via adjusted R^2 and MSE; assessed robustness using Cook's distance
- Glassdoor Data Science Jobs Analysis *July 2023*
  - Scraped 1K+ job listings; analyzed skill demand and salary variation using pandas, seaborn, and NLP text mining
  - Applied TF–IDF and clustering to extract trending keywords and regional hiring patterns
- Rock vs Mine Prediction *July 2023*
  - Classified sonar readings using Logistic Regression and SVM achieving 85% accuracy
  - Performed correlation and ANOVA to identify statistically significant signal attributes
- Analytics Mini Projects (Hiring, Loan, Calls, IMDB, Games, YouTube, Instagram) *June 2023*
  - Conducted EDA, regression, and classification analyses on diverse datasets using scikit-learn, pandas, and NumPy
  - Applied clustering, hypothesis testing, and feature selection to improve insight precision across 7+ datasets
- XYZ-ADS (Ad Analytics) *June 2023*
  - Created Python dashboards tracking CTR, conversion rate, and ROI metrics for ad-campaign performance
  - Applied multiple regression and chi-square tests to assess ad-format effectiveness
- Water Quality Analysis using Satellite Images *Aug 2022 – May 2023*
  - Used Random Forest and SVM to classify turbidity levels from multispectral satellite imagery
  - Integrated geospatial layers with regression outputs; achieved 88% accuracy visualized via folium maps

Software Engineering, Cloud & Systems Design

- AWS 3-Tier Architecture Project *Oct 2025*
  - Deployed secure 3-tier web architecture using AWS EC2, RDS, and S3 with load balancing and IAM access control
  - Integrated CloudWatch monitoring, CI/CD pipelines, and auto-scaling to validate high availability under load
- Soccer Web App (Advanced Database Systems Design) *Aug 2023 – Dec 2023*
  - Built Bottle (Python) CRUD web app with normalized SQLite schema and foreign-key relationships for team data
  - Optimized SQL queries achieving <200ms latency; implemented search and update endpoints for user operations

Information Visualization & Dashboards

- Interactive Grocery Store Visualization System (D3.js) *Jan 2025 – May 2025*
  - Created multi-view D3.js dashboard (bar, donut, chord) analyzing demographic spending and income-based preferences
  - Implemented brushing, hover tooltips, and transitions in JavaScript ES6; deployed via Kent State web server

Simulation & Modeling

- Infection Simulation (Mesa) *July 2025*
  - Modeled epidemic spread using agent-based simulation in Python (Mesa) with adjustable R_0 and recovery parameters
  - Visualized contagion curves and intervention effects through real-time grid animation and matplotlib plots

Sports Analytics

- Premier League Analytics Suite (2000–2023) *Jan 2023*
  - Analyzed 22 seasons of Premier League data for team efficiency using pandas, seaborn, and matplotlib
  - Built visual models for shot maps, expected goals (xG), and seasonal regressions of win probability

Certifications

- AWS Cloud Technical Essentials – Amazon Web Services Oct 2025
- Biology Meets Programming, UC-San Diego(UCSD) Aug 2025
- AWS Educate – Introduction to Generative AI Aug 2025
- Building Computer Vision Applications with Python Jul 2025
- Data Analysis with R Programming (Google) Nov 2024
- GenAI with Diffusion Models (NVIDIA) Oct 2024
- AWS Academy Graduate – Cloud Foundations May 2023
- AWS Academy Graduate – Data Analytics May 2023
- ITIL-aligned Data Analytics Virtual Internship (AICTE NEAT) May 2023