Praveen Puviindran

praveen.puviindran@gmail.com • GitHub • Wellesley Hills, MA • LinkedIn • Personal Website

EDUCATION

University of North Carolina at Chapel Hill **May 2025**
B.S. Statistics and Analytics *Chapel Hill, NC*

WORK EXPERIENCE

National Institutes of Health | Andrew D. Johnson, Ph.D. **Sep 2025 – Present**
Post-baccalaureate Research Fellow *Framingham, MA*

- Developing Python pipelines analyzing 3K proteins (N=25K), identify biomarkers for heavy menstrual bleeding.
- Applying pathway enrichment and protein-network analyses to uncover causal mechanisms driving HMB.
- Building LASSO-based proteomic risk models to construct diagnostic panels and reduce detection latency.

UNC School of Medicine, Dept. of Virology| Dirk Dittmer, Ph.D. **Jan 2025 - May 2025**
Data Science Consultant *Chapel Hill, NC*

- Engineered a q-PCR-ready classifier in Python using viral RNA-seq for a clinical solution for HIV+ patients.
- Applied PCA (**silhouette = 0.77**) to synthesize tumor profiles, informing subtype-specific treatment options.
- Led model selection/validation process (LASSO, SVM, RF, MLP, **AUC ≥ 0.95**), used SHAP for interpretability.

MD Anderson Cancer Center, Dept. of Radiation Oncology | Lauren Colbert, MD **Jun 2024 - Aug 2024**
CPRIT/CURE Summer Research Bioinformatics Intern *Houston, TX*

- Built microbiome data pipelines (QIIME2, R) tracking radiation-induced gut shifts in cancer patients.
- Applied hypothesis-driven testing (PERMANOVA) to find microbiome shifts linked to skin toxicity risk.

DATA SCIENCE PROJECTS

NBA Synergy Engine - Deep Learning Roster Optimization System **Nov 2025**

- Built an ML pipeline o 170K+ possessions, migrating tracking data into a normalized SQLite database.
- Trained a permutation-invariant DeepSet model predicting lineup synergy from GMM archetypes (40 RMSE).
- Designed a vectorized Generative GM module simulating 450+ roster options to choose the optimal 5th starter.

Early Diabetes Risk Prediction **Nov 2024**

- Queried 253k+ health records (SQL), engineered features, trained Random Forest (SMOTE, **85% accuracy**).
- Used SHAP-based feature analysis to identify high-risk patterns and refine early-screening criteria.

NBA Game Outcome Forecasting **April 2024**

- Built Random Forest models on 1,100+ games ($R^2 = 0.84$) using real-time API data to forecast game outcomes.
- Simulated alternative outcomes using historical data to explore optimized game plans and flag outlier strategies.

LEADERSHIP & EXTRACURRICULARS

UNC Musical Empowerment — Senior Co-President **Aug 2021 - May 2025**

- Led 140+ volunteers across 13 chapters; managed operations, training, and logistics for community programs.

North Carolinians Youth for Peace — Founder & Lead Instructor **Mar 2021 - Sep 2024**

- Built a virtual ESL program for 40+ Sri Lankan students; created curriculum, led a 6-person international team.

TECHNICAL SKILLS

- **Languages:** Python (pandas, NumPy, scikit-learn, PyTorch), R, SQL, MATLAB, Git
- **Machine Learning:** Forecasting, supervised learning, classification, model interpretability (SHAP)
- **ML & Analytics:** Forecasting, supervised learning, linear/integer programming, EDA, causal inference
- **Data Pipelines:** ETL development, reproducible modeling workflows, API integration, version control