Praveen Puviindran

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

Data Scientist specializing in high-dimensional biomedical modeling and applied machine learning systems. Experienced in building end-to-end predictive workflows across large-scale biomedical and sports datasets.

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*

Proteomic Biomarker Modeling for Heavy Menstrual Bleeding

- Performed PWAS across 2,922 proteins (315 cases, 26K+ controls), identifying 34 significant biomarkers.
- Built proteomic risk score achieving AUC = 0.891; integrated with clinical+genetic features (**AUC = 0.925**).
- Validated age-stratified performance (**AUC ≈ 0.83**), confirming consistent predictive accuracy across subgroups.

Machine Learning Classification of Platelet Reactivity Extremes

- Built supervised pipeline on 2,680 participants to classify top/bottom 10% platelet reactivity extremes.
- Achieved **AUC 0.842** with gradient boosting against random forest and logistic regression with leakage controls.
- Applied median imputation and scaling; feature importance identified CRP and fibrinogen as top predictors.

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.

DATA SCIENCE PROJECTS

NBA Synergy Engine - Deep Learning Lineup Optimization System **Feb 2026**

- Designed SQL-backed feature pipeline across 170K+ possessions modeling lineup-level interaction effects.
- Applied PCA and Gaussian Mixture Models to define modern NBA archetypes from high-dimensional metrics.
- Implemented permutation-invariant DeepSet neural network predicting five-player synergy (**RMSE ~40**).
- Built vectorized simulation module evaluating 450+ roster combinations to rank optimal fifth-player fits.

NBA ShotIQ - Probabilistic Shot Quality and Efficiency Engine **Jan 2026**

- Developed expected shot value model estimating P(make | location, context) from NBA shot chart data.
- Evaluated probabilistic performance using log loss, Brier score, and calibration curves to ensure reliability.
- Engineered player-level metrics (SMOE, Shot Diet Difficulty) separating shot selection from shot-making skill.
- Built interactive visualization layer generating half-court frequency and over/under-performance heatmaps.

Early Diabetes Risk Prediction **Nov 2024**

- Queried and filtered 253k+ BRFSS health records using SQL to extract target cohorts and health features.
- Cleaned and engineered features in Python; applied SMOTE to address class imbalance and improve models
- Trained Random Forest models (**85% accuracy**); optimized with cross-validation and hyperparameter tuning.

TECHNICAL SKILLS

- **Languages:** Python (pandas, NumPy, scikit-learn, PyTorch), R (tidyverse, caret, glmnet), SQL
- **Machine Learning:** classification, neural networks, supervised learning, cross-validation, SHAP
- **Data & Databases:** PostgreSQL, ETL development, API integration, schema design, version control
- **Cloud & Deployment:** AWS (S3), Streamlit, Git, GitHub, Docker