

Prithika Roy

✉ prithika_roy@berkeley.edu ☎ (919) 454-5081 🔗 LinkedIn

Education

University of California, Berkeley - Berkeley, CA

Expected Graduation: May 2031

Incoming PhD Student in Statistics - Supported by Chancellor's Fellowship

- **Research Interests:** Interpretable machine learning, uncertainty quantification, statistical inference

University of North Carolina at Chapel Hill - Chapel Hill, NC

May 2026

B.S. in Statistics & Analytics, B.A. in Computer Science | Minor in Business Administration | GPA: 3.99

- **Graduation with Highest Honors in Statistics & Analytics**

Patient Length-of-Stay Prediction and Inpatient Bed Forecasting in UNC Hospitals' Emergency Department

Advisors: Dr. Chudi Zhong & Dr. Serhan Ziya

- **Relevant Coursework:** Optimization, Probability, Real Analysis, Stochastic Modeling, Multivariable Calculus, Methods of Data Analysis, Linear Algebra, Machine Learning, Deep Learning, Biostatistics, Simulation for Analytics
- **Activities:** Member, Women in Computer Science; Product Manager & **Centible**  Team Co-Lead, App Team Carolina; Grader for STOR 235: Mathematics for Data Science
- **Honors & Awards:**
 - Statistics & Analytics Award (2026), UNC Department of Statistics and Operations Research
 - Grace Hopper Conference Scholarship (2025), UNC Department of Computer Science
 - National Merit Scholarship (2022), National Merit Scholarship Corporation
 - Ricky Hendrick Scholarship (2022), Hendrick Automotive Group
 - President's Volunteer Service Award (2022), President's Council on Service and Civic Participation

Research Experience

UNC Department of Statistics and Operations Research - Chapel Hill, NC

Aug 2024 – Present

Undergraduate Researcher

- **Clustering and Summarizing the Rashomon Set of Sparse Decision Trees**  (Advisor: Dr. Chudi Zhong)
 - Developed vector representations of decision trees and applied clustering and dimensionality reduction techniques (e.g. t-SNE, PaCMAP, LocalMAP) in Python to efficiently sample representative trees from the Rashomon set of well-performing models
 - Quantified predictive redundancy arising from shared decision boundaries within the Rashomon set. Evaluated a new ensemble construction method that utilizes trees selected from distinct clusters of similar models, benchmarking its performance against the optimal model and ensembles of randomly-selected trees
- Senior Honors Thesis (Advisors: Dr. Chudi Zhong & Dr. Serhan Ziya)
 - Applying machine learning methods (quantile regression, random forest, etc.) to produce forecasts of patient length-of-stay in the emergency department using clinical and operational data from UNC Hospitals
 - Developing a framework to aggregate admission probability and length-of-stay model outputs to forecast inpatient bed demand; incorporating conformal prediction methods to generate prediction intervals with approximately valid marginal coverage

NC Institute for Climate Studies - Asheville, NC

Jul 2023 – Dec 2023

Statistical Research Intern

- Developed scripts using Python to analyze environmental data and explore relationship between fractional available water (FAW) and wildfire development to explain and predict fires
- Processed and cleaned >700 GIS soil moisture raster datasets & estimated measures of FAW using NASA surface soil moisture data and wilting point/field capacity metrics
- Created time series of FAW conditions prior to major wildfire events by computing averages over burned area maps

UNC Lineberger Cancer Center - Chapel Hill, NC

Aug 2022 – Dec 2022

Undergraduate Research Assistant for Internet Tobacco Vendors Study

- Collected and cleaned data on online tobacco retailers' transparency practices for quantitative analysis informing statewide tobacco regulation and health policy
- Trained in survey methodology and statistical analysis in SAS through weekly research team sessions

Professional Experience

Oliver Wyman - Raleigh, NC

June 2025 – Aug 2025

Quantitative Analyst Intern

- Developed key components of a credit risk rating model in Python to estimate probability of default for an equity derivatives clearinghouse. Implemented structured imputation and variable transformation logic that improved model performance and data utilization, along with functionality for exploratory data analysis and model fitting/selection
- Designed modular testing code to validate modeling assumptions, perform sensitivity analyses and assess stability of parameters, and benchmark outputs against external credit ratings
- Analyzed regulatory data to establish the model's default definition. Authored 60+ pages of documentation detailing data processing & statistical methodology to ensure defensibility of modeling decisions; built an interactive Excel tool to simulate credit risk predictions, enabling client to execute model on custom inputs to explore its mechanics

WillowTree (TELUS Digital) - Charlottesville, VA

May 2024 – Aug 2024

Analytics Architect Intern

- Implemented site tracking and audited existing analytics infrastructure, facilitating data-driven marketing for an international supermarket chain. Authored comprehensive documentation and translated analytics feature requirements into clear, testable tickets for software engineers
- Constructed SQL queries to conduct user path analysis and calculate month-to-month revenue changes, assessing the impact of exclusive product advertisements on overall sales for a state-run alcohol agency
- Evaluated current analytics architecture, proposed key metrics and improvements for accurate data collection, and developed Google Analytics user and session traffic acquisition reports for a major Virginia healthcare system's websites

The Daily Tar Heel - Chapel Hill, NC

Jan 2023 – Present

Senior Data Reporter

- Collaborate with writers to create data visualizations for articles using Datawrapper, Flourish, R, and Tableau
- Produced Lumbee Tribe member population density map of NC by county [📄](#), visualized variations in UNC women's basketball attendance over time [📄](#), analyzed maternal mortality disparities in NC by race [📄](#), and mapped residence halls containing asbestos [📄](#)

Technical Skills

- **Programming Languages:** Python, Java, C/C++, SQL, R, SAS, MATLAB
- **Tools:** GitHub, Tableau, Google Analytics, Adobe Analytics (Business Practitioner Certification)

Relevant Projects

Bitcoin Price Forecasting Using Prediction Markets: [📄](#) Trained and evaluated LSTM, XGBoost, and transformer-based models in Python, leveraging Kalshi prediction market data and historical Bitcoin prices to enhance short-term Bitcoin price forecasting