

# SOUVIK PAUL

Chicago, IL-60612

Email: [spaul33@uic.edu](mailto:spaul33@uic.edu)

Phone: +1 872 258 4413

[LinkedIn](#)

---

## Education

---

Aug 2022 – Present: **University of Illinois Chicago (UIC)** Chicago, USA

Ph.D. in Biostatistics (GPA: 4.00)

**Courses:** Linear Models, Generalized Linear Models, Statistical Inference, Longitudinal Analysis, Generalized Mixed Effect Models, Computational Statistics

Jul 2019-May 2021: **Indian Institute of Technology Kanpur** Kanpur, India

MS in Statistics (GPA: 3.60)

**Key Project:** Comparison of MCMC samplers Random Walk Metropolis Hastings, Metropolis Adjusted Langevin Algorithm and Barker

**Achievement:** Top 1 percentile in the entrance exam for Master's in Indian Institute Technology

Jul 2016-May 2019: **University of Calcutta (RKMRC)** Kolkata, India

BS in Statistics (GPA: 3.52)

**Achievement:** (i) Recipient of prestigious INSPIRE Scholarship by Department of Science and Technology, Govt. of India

(ii) Placed top 3 in Department of Statistics, University of Calcutta

## Skills

---

R, Python, R Shiny, SAS, Pymc (Bayesian), STATA, Excel

## Professional Experience

---

May 2023-Present: **Research Assistant**

Center for Clinical and Translational Science, University of Illinois Chicago

- Work as Biostatistics consultant – statistical analysis for the clients come for, mostly, health data.

May 2023-Present: **Graduate Research Assistant**

Dr Supriya Mehta (Rush University), Dr Runa Bhaumik (UIC)

- Work on research project, in Kenya, on adolescent girls. A longitudinal study on BV, HSV, STI, HIV with taxa, mostly using Mixed effect models.

Aug 2022-May 2023: **Graduate Teaching Assistant**

University of Illinois Chicago

- Tutored the Graduate Courses ``Biostatistics-I'' and ``Health Analytics using Python''

May 2021-Jul 2022: **Data Scientist, Machine Learning Division**

OpsCruise Inc., Chennai, India (A US Based Company)

- Machine Learning model development and validation, related to microservices and containers data and data segregation from normal and anomalous traffic using unsupervised models.
- Employed Nelder-mead and Genetic optimization Algorithm for Hdbscan clustering model, and tested artificial database with mixed-type features using AWS Densclus package

- Worked on Convergent Cross Mapping (CCM) - a causal inference procedure to identify the causal association over the correlation.
- Had weekly meetings with C-level and with outside vendors to solve the in-time issues in the container system. Reduced 90 % False alerts in the customer database by identifying significant features

## Selected Course Projects

---

Dec 2022-Jan 2023: **Correlation Behavior Before and After Discretization**

University of Illinois Chicago, USA

- A simulation study on all pairs (based on skewness) of variables. Discretized the variables from 2 to 100 levels based on quantiles. Key observations are- (i) in one case, correlation increases after discretization, (ii) in other case, correlation increases after discretization when given correlation is high and decreases, for the same pair, when given correlation is low.

Jan 2021-Apr 2021: **Variable Selection in Non-Parametric Regression**

Indian Institute of Technology Kanpur, India

- Used Nadaraya Watson Kernel Estimator and Gasser-Muller Estimator for non-parametric regression. Firstly, selected best set of variables having minimum MSE from all possible combinations of variables and secondly, using backward elimination.

Oct 2020-Dec 2020: **Time Series Analysis and Forecasting on Minimum Temperature Data**

Indian Institute of Technology Kanpur, India

- Exploratory data analysis and cleaning including missing value imputation, randomness analysis and trend analysis.
- Performed seasonality analysis, used augmented Dickey-Fuller (ADF) test for stationarity and fit the ARMA accordingly.

May 2020-Jul 2020: **A Study on Facebook Metrics Data - A Linear Model Project**

Indian Institute of Technology Kanpur, India

- Executed exploratory data analysis for advertising impact research, including data visualization, missing data Imputation, influential points detection, normality validation.
- Performed linear model with box-cox transformation and dealt with the issue of heteroscedasticity and multicollinearity. Employed regularization skills (such as Lasso) to select essential features for the purpose of prediction.

Apr 2020-Jul 2020: **Wisconsin Breast Cancer Classification**

Indian Institute of Technology Kanpur, India

- Exploratory data analysis and visualization including box plot, outlier detection and missing value imputation.
- KNN, Logistic Regression, LDA, Decision Tree, Random Forest classifiers were executed and Random Forest came out to be the best with both Recall and F1-score 98 %. Feature Selection using Random Forest and top 3 (out of 12) produced 97 % Recall.

## Selected Courses

---

Time Series Analysis, Machine Learning, Bayesian Methods, MCMC, Survival Analysis

## Volunteer Experiences

---

- Work in Rescuing Leftover Cuisine, Chicago, USA  
Work as a volunteer to distribute the leftover foods from restaurant to the community in need
- Leader of National Service Scheme India in University of Calcutta  
Guided around 30 students and worked with them for social services.