Rahul Raoniar

Researcher, Data Science and Machine Learning Enthusiast.

+91-9971705132 rahul.raoniar@outlook.com

LinkedIn Blog GitHub Medium YouTube ResearchGate

EDUCATION

Indian Institute of Technology Guwahati (Assam) — **July 2016 – Present**
Ph.D. in Transportation Systems Engineering — *Thesis Submitted*

AcSIR - Central Road Research Institute (CSIR Lab), New Delhi — **May 2013 – May 2015**
Master of Technology in Transportation Engineering — *C.G.P.A. - 9.10*

Teerthankar Mahaveer University, Moradabad (Uttar Pradesh) — **May 2009 – May 2012**
Bachelor of Technology in Civil Engineering — *C.G.P.A - 8.21*

Raiganj Polytechnique, Raiganj (West Bengal) — **May 2006 – May 2009**
Diploma in Civil Engineering — *Percentage - 77.8%*

PROFESSIONAL EXPERIENCE

YouTube Content Creator and Blog Writer — **July 2018 – Present**
Data Science Content Creator — *India*

- Created practical content about data science for students and educators.
- Published over 50+ videos and 40+ blogs on data science tutorials.

Road Safety Researcher — **July 2016 – Present**
Ph.D. Research Scholar — *IIT Guwahati, India*

- **Thesis Title**: Study of Pedestrians' Unsafe Road Crossing Behaviour at Signalised Intersection Crosswalks.
- Developed traditional and advanced statistical models to quantify pedestrian risk-taking behaviour at signalised intersection crosswalks.

Trainee Scientist, CSIR-Central Road Research Institute Lab. — **May 2013 and May 2015**
Trainee Scientist — *New Delhi, India*

- Evaluated the qualitative factors that diminished or increased system usage for both bus and train users for Delhi city.
- Analyzed transportation performance using questionnaire survey and Structural Equation Modelling (SEM) technique.

TECHNICAL SKILLS

- **Programming Languages:** Python, R and MySQL.
- **Core Transportation:** Road Users' Safety Evaluation, Vulnerable Road Users, Injury Epidemiology, and Injury Prevention, Transportation Planning and Traffic Engineering.
- **Applied Data Science:** Data Science Pipeline (Data collection, Manipulation, Exploratory analysis, Statistical Analysis, and Report making), Statistics (Experimental design, Exploratory and Confirmatory data analysis, Hypothesis testing, and Bayesian A/B testing), Geospatial data analysis and Visualisation, Time series forecasting, Big data analytics, OOP, Git and GitHub, Google sheet, and Excel.
- **Applied ML:** **Classification** (Binary, multi-nominal and ordinal logit models) and **Regression** (Multiple linear regression, lasso and ridge regression), **Count models** (Poisson, negative binomial, and zero-inflated regression), **Clustering** (k-means, hierarchical), **Survival analysis** (KM estimate, COX-PH and AFT models), **Mixed Effects Models** (random intercepts and slopes), **Time Series Forecasting**, Association rule mining, Decision trees, **Ensemble models** (bagging and boosting) and **Deep Learning**.
- **Core/Analytics Tools:** Linux CLI, Stata, Tableau, QGIS, LaTeX, MS Office Suite, Notebooks (Jupyter, Google collab etc.) and IDEs (VS Code and PyCharm).
- **Searching:** Googling and Searching Stack Overflow.
- **Libraries/Packages:** Data Manipulation (pandas, dplyr and dfply), Visualization (ggplot2, matplotlib, seaborn, plotnine, altair and plotly), Feature Engineering and Selection (sklearn and feature_engine), Geospatial Analysis and Visualisation (folium and geopandas), Dashboard (plotly Dash and tableau), Text Analysis (regex), Time Series Forecasting (prophet and statsmodels), Machine Learning (scikit learn, statsmodels, tidymodels, pycaret, keras, tensorflow and H2o), Big Data Analytics (PySpark) and Web Application (streamlit and flask).

PROJECTS

Ph.D. Projects | *IIT Guwahati, India*

- Formulated plan, collaborated with eight students and collected video graphic data of pedestrians' road crossing behaviour from 11 intersections across Kolkata city.
- Designed distraction-themed questionnaire, estimated required sample size, trained interviewers for the conversation-styled interview, conducted face-to-face interviews across Kolkata city and identified factors influencing pedestrian's distracted road crossing behaviour using binary logistic regression models.
- Analysed the parents' role in school mode choice for their children in Guwahati city using multinominal logit model. The study highlights the role of parents in the mode choice process through their perceptions of safety, economic standards, and child characteristics. {}
- Analysed pedestrian foot-over bridge utilisation across four Indian cities (14 locations) using tree-based ensemble techniques. {}
- Identified social and non-social factors influencing pedestrian's signal violation behavior at intersection crosswalks using binary logistic regression model. {}
- Identified the optimal waiting time of pedestrians at intersection crosswalks using survival analysis (COX Proportional Hazard and Accelerated Failure Time models). The study results could be used to propose optimal red-phase length for pedestrian signals at crosswalks. {}

Data Science and Machine Learning Projects | *Tools: Python, R, Stata, MySQL & Tableau*

- Trained, dockerized, and deployed a house-price prediction model using GitHub actions on the Heroku cloud. {}
- Trained, dockerized, deployed and tested a churn prediction model on a local Linux machine. {}
- Built a Streamlit-based probability distribution fitter web application and deployed it on Heroku cloud. The application compares 80+ probability distributions and ranks them based on their fit to the data. {}
- Built a Streamlit-based early diabetes prediction web application and deployed it on Heroku cloud. {}
- Forecasted Seattle Fremont Bridge Daily Bicycle Counts using Meta's Prophet Python library. {}
- Predicted lung cancer patients' survival using Cox Proportional Hazard model (survival analysis). {}
- Built a concrete strength prediction regression model explainer using Shapley values {} and tuned its hyper-parameters using Grid, Random and Genetic-based search using scikit-learn library {}. The Shapley value method is utilised to identify the contribution of each variable in the model and its direction of influence (+ve or -ve).
- Built choropleth maps using Python (folium and geopandas) and QGIS. {Map 1 Map 2 Map 3}
- Built a parking space counter using OpenCV library. The application counts the number of empty and occupied spaces from a video feed of a parking lot.
- Simulated email link click-through rate using Bayesian A/B testing. A Monte Carlo simulation was performed to identify which email variant performs better and to what extent.
- Optimised deep Convolutional Neural Network model's hyper-parameters using Keras and sklearn libraries. {}
- Achieved 98.97% classification test accuracy MNIST digit classification using PyTorch. {}
- Created Sales, HR and Accounts Key Performance Indicators (KPIs) dashboard using Tableau and MySQL. {}

LEADERSHIP

Developer of "pysustrans" Python library | *May 2022 - Present*

- Developed Open-Source data analytics library for performing regular analytics tasks specific to transportation data.

Taught "Python for Scientific Computing and Data Science" | *April 2022, IIT Guwahati, India*

- Taught "Python for Scientific Computing and Data Science" in a three week long workshop with 300+ registered participants.

Guided Master's Students [MTP Project] | *2018-Present, IIT Guwahati, India*

- Collaborated with Ph.D. guide and four Master's students to formulate hypothesis, perform data analysis, interpret results and prepare dissertation report.

Teaching Assistant in Transportation Planning and Traffic Engineering Lab. | *2016-2022, IIT Guwahati, India*

- Collabortated with instructor and 6 other TAs to lead recitations, grade labwork, and answer 60+ students' queries.

AWARDS AND ACHIEVEMENTS

- Honoured by university for the top academic performance in the department of Civil Engineering.
- Got selected as Trainee Scientist in CSIR-CRRI lab among top 1% candidates.
- Received full scholarship from CSIR for Master's and MHRD for Ph.D. education.

COURSES, TRAININGS, AND CERTIFICATIONS

- Data Scientist with Python Track (DataCamp Specialization)
- Data Scientist with R Track (DataCamp Specialization)
- Machine Learning Scientist with Python Track (DataCamp Specialization)
- Statistical Analysis with R for Public Health (Coursera Specialization)
- AI for Medicine (Coursera Specialization)
- Deeplearning.Ai TensorFlow Developer (Coursera Specialization)
- MySQL Boot Camp (Udemy)
- Tableau Fundamental Track (DataCamp Specialization)
- Road Safety Auditor Certification (MoRTH and IRC)

PUBLICATIONS

Journals **2015 - Present**

Published

- ***Raoniar, R.**, Maqbool, S., Pathak, A., Chugh, M. and Maurya, A. K.*, 2022. Hazard-Based Duration Approach for Understanding Pedestrian Crossing Risk Exposure at Signalised Intersection Crosswalks - A Case Study of Kolkata, India, Transportation Research Part-F: Traffic Psychology and Behaviour, vol. 85, pp. 47-68.
- ***Raoniar, R.** and Maurya, A. K., 2022.* Pedestrian Red-Light Violation at Signalized Intersection Crosswalks: Influence of Social and Non-Social Factors, Safety Science, vol. 147.
- ***Raoniar, R.** and Maurya, A. K.*, 2021. Pedestrian Crossing Behaviour at Signalized Intersection Crosswalks: An Observational Study of Factors Influencing Pedestrian Walking Speed, Safety Margin, and Violation, Journal of the Eastern Asia Society for Transportation Studies, vol. 13.
- *Das, S., Boruah, A., Banerjee, A., **Raoniar, R.**, Nama, S. and Maurya, A.K.* 2021. Impact of COVID-19: A Radical Modal Shift from Public to Private Transport Mode, Transport Policy, vol. 109.
- *Banerjee, A., **Raoniar, R.** and Maurya, A.K.*, 2020. Pedestrian Overpass Utilization Modelling Based on Mobility Friction, Safety and Security, and Connectivity Using Machine Learning Techniques. Soft Computing, vol. 24, 17467–17493.
- ***Raoniar, R.**, Das. T, Banerjee, A. and Maurya, A.K.*, 2019. The Parents' Role in School Mode Choice for their Children: A Case Study in Guwahati, Journal of the Eastern Asia Society for Transportation Studies, vol. 13, pp. 775-94.

Conferences **2015 - Present**

Presented

- *Banerjee, A., **Raoniar, R.** and Maurya, A. K.*, 2021. Study of Factors Impacting Safety-Security and Mobility Friction on the Choice of Pedestrians in Using Skywalk Facilities Through Soft Computing Approaches, Journal of the Eastern Asia Society for Transportation Studies, vol. 13.
- ***Raoniar, R.** and Maurya, A.K.*, 2021. Hazard-Based Duration Approach: Pedestrian Crossing Risk Exposure at Signalised Intersections in Kolkata, India, 6^{th} Conference of the Transportation Research Group of India, ID-24, Tiruchirappalli, India.
- ***Raoniar, R.** and Maurya, A.K.*, 2021. Impact of Different Distractions on Pedestrian Road Crossing Behaviour at Signalized Intersection Crosswalks, [ID: TRBAM-21-01200 (Poster presentation)], 100th Annual Meeting of Transportation Research Board, Washington, DC, United States.
- ***Raoniar, R.**, Maqbool, S. and Maurya, A.K.*, 2020. Factors Influencing Pedestrian Walking Speed and Safety Margin at Signalized Intersection Crosswalks, 13th TPMDC, Paper ID 86, IIT Bombay.
- *Banerjee, A., **Raoniar, R.** and Maurya, A.K.*, 2020. Identification of Factors Influencing the Pedestrian Walking Speed Over Elevated Facilities Using Deep Neural Network, 13th TPMDC, Paper ID 136, IIT Bombay.
- ***Raoniar, R.** and Maurya, A.K.*, 2020. Predicting Pedestrian's Red-Light Violation (RLV) Behavior at Signalized Intersection Crosswalks using Machine Learning Techniques, ASCEic (ID: AIC2020-10-932).