

NAMAN BAREJA

(646) 200-1041 | nb3236@columbia.edu | <https://www.linkedin.com/in/namanbareja/> | <https://github.com/SodiumMan91>

EDUCATION

Columbia University

Dec 2025

Master of Science, Data Science, GPA: 3.63/4

Leadership: President, Data Science Institute Student Council, Lead Department Representative, TA for Deep Learning for NLP

Coursework: Statistical Inference and Modeling, Machine Learning for Finance and Functional Genomics

Delhi Technological University (formerly Delhi College of Engineering)

May 2023

Bachelor of Technology, Mathematics and Computing, GPA: 8.92/10

TECHNICAL SKILLS

Languages: Python, R, SQL, SAS, MATLAB

ML & Data: scikit-learn, XGBoost, PyTorch, TensorFlow, NLP (NLTK, Gensim), Pandas, Numpy

Tools: AWS, BigQuery, Tableau, Power BI

EXPERIENCE

Bendesky Lab, Zukerman Institute

Sep 2025 – Dec 2025

Research Assistant

- Built computer vision pipeline using YOLO to detect and track real vs. 3D-printed fish frame-by-frame, enabling behavioral quantification at scale
- Trained action classification model (biting, flaring) to quantify aggression behaviors across experiments

Mailman School of Public Health, Department of Epidemiology

May 2025 – Dec 2025

Machine Learning Research Assistant

- Built hierarchical ensemble NLP pipeline to classify sentiment in Reddit posts on polysubstance use, combining lexicon and transformer-based models
- Improved classification accuracy by 23% by aggregating sentence-level sentiment within contextual windows, outperforming post-level baselines

Microsoft

Mar 2025 - May 2025

CX Research Intern

- Analyzed internal documentation and industry benchmarks to identify AI applications in customer research and cybersecurity, delivered biweekly strategic reports to cross-functional teams
- Proposed AI-driven evaluation metrics and research frameworks, delivering forward-looking investment recommendations

EXL Services

Nov 2023 - Aug 2024

Business Data Analyst

- Re-engineered credit risk model pipelines (CCAR, CECL) using SQL, improving data integrity across 500K+ financial records
- Automated and optimized model pipelines in SAS, reducing processing time by 15% and minimizing manual errors; collaborated with model owners to validate and deploy enhancements
- Improved Probability of Default (PD) model accuracy by 10% through feature diagnostics and statistical analysis; presented findings to senior leadership

IFFCO Tokio General Insurance Co. Ltd.

May 2022 - Jul 2022

Machine Learning Engineer Intern

- Designed and implemented a computer vision pipeline for automated damage detection and repair cost estimation in car insurance claims using Mask R-CNN, InceptionV3, and manual annotations from VGG Image Annotator, improving processing efficiency and reducing processing times from 3 weeks to 3 days for claims of small monetary value
- Developed a customer retention prediction model using XGBoost and regression improving customer retention by 7%

Centre for Artificial Intelligence and Robotics, DRDO

May 2021 - Sep 2021

Student Research Trainee

- Researched extension of semi-supervised Snowball model for bi-directional related components extraction from plain text by leveraging 3 seed instances of connectivity relationship for label propagation across 10K+ sentences using NLTK and Gensim
- Boosted F1-score by 15% over unsupervised baselines through ensemble modeling and hyperparameter tuning via successive halving
- Compared two clustering methods, analyzing trade-offs in scalability and label efficiency; presented findings at NCETET 2023

PROJECT

Predictive Analysis of Poor Mental Health Prevalence across neighborhoods

Sep 2024 - Dec 2024

- Led team of five to model neighborhood-level mental health prevalence using demographic and environmental data
- Built ML models (XGBoost, scikit-learn) achieving 78% F1-score; identified 10+ key predictors to inform targeted public health interventions