

Sagnik Chakravarty

College Park, MD | sagnikch@umd.edu | (240) 610-7356 | github.com/Sagnik-Chakravarty | linkedin.com/in/sagnik-chakravarty | scholar.google.com/citations?user=thxuXDwAAAAJ&hl=en

EDUCATION

University of Maryland, College Park, USA- MS in Survey and Data Science | **GPA:**3.653 August 2024 - May 2026

- **Coursework:** Machine Learning I & II (SURV615/616), Fundamentals of Data Collection I & II (SURV621/622), Sampling I (SURV627), Long-Context Language Models (CMSC8480), Data Visualization (SURV727), Statistical Methods, Causal Inference & Experimental Design (SURV627), Research Ethics (SURV612).

Vellore Institute of Technology, Chennai, India- MS in Data Science | **GPA:**3.575 September 2022 - July 2024

- Graduated with First Class and Distinction, Best Student Researcher Award
- **Coursework:** Machine Learning, Deep Learning, Statistical Natural Language Processing, Multivariate Data Analysis, Regression & Predictive Modeling, Time Series Analysis, Artificial Intelligence, Neural Networks & Fuzzy Logic, Optimization Techniques, Programming for Data Analysis, Statistical Inference, Topological Data Analysis (via project/research).

Amity University, Kolkata, India- BS in Statistics | **GPA:**3.637 August 2018 - July 2021

- Graduated with First Class and Distinction
- **Coursework:** Probability Theory I & II, Statistical Inference I & II, Sampling Theory, Linear Models, Design of Experiments, Time Series Analysis, Stochastic Processes, Nonparametric Statistics, Applied Statistics, Statistical Computing (C Programming).

RESEARCH EXPERIENCE

Research Assistant, University of Maryland, College Park- Research Assistant September 2025 - *Present*

- Collected and preprocessed large-scale social media datasets (Reddit, YouTube, Bluesky etc), newspaper, magazines etc to study public discourse on artificial intelligence using natural language processing (NLP) methods.
- Developed computational models for frame detection and clustering with large language models, improving the interpretability of online discourse analysis.
- Conducted survey experiments with nationally representative samples to assess the effects of AI framing on public attitudes, trust, and behavioral intentions.
- Collaborated with the PI on analysis, interpretation, and preparation of findings for publication and conference submission, integrating political communication theory with computational methods.

Gard Lab, University of Maryland, College Park- Research Assistant February 2025 - September 2025

- Analyzed psychological survey data to study cognitive processes and behavioral patterns using statistical techniques such as ANOVA, t-tests, and regression in R.
- Automate stimulus workflow in PsychoPy from previous JavaScript and HTML code, improving precision in stimulus presentation, response tracking, and experimental timing by 10%. Develop a data pipeline to process, store, and visualize participant test results, integrating survey responses and behavioral data into an interactive dashboard for real-time feedback and analysis.

Strickland Law Firm, Remote- Survey Data Analyst December 2024 - February 2025

- Led a team of three to analyze 2 million data points from Reddit, X, Instagram, Facebook, news sources, and music platforms to measure shifts in public sentiment toward the YSL Rico Trial due to tweets from ThuggerDaily
- Utilized Playwright for large-scale web scraping, conducted sentiment analysis with NLTK and TextBlob, applied LDA for topic modeling, and performed causal inference using DAGs (R's dagitty package), uncovering significant sentiment shifts

Indian Statistical Institute, Kolkata- Data Science Research Intern May 2023 - June 2024

- Developed a Graph Convolutional Neural Network (PyTorch-Geometric, nilearn) using FMRI data to detect early-stage Alzheimer's (75% accuracy) and Parkinson's (72% accuracy)
- Simulated C. elegans locomotion using the NEURON package, implementing LSTM (TensorFlow/Keras) and GRU (PyTorch), with LSTM outperforming GRU by 9% in neuron identification accuracy
- Designed an optical illusion detection pipeline using ConvLSTM (TensorFlow) and Nvidia FlowNet2, preprocessing 10M+ frames with OpenCV, achieving 85% accuracy in detecting visual illusions

Clarkson University, Remote- Research Data Science Intern September 2023 - January 2024

- Collaborated remotely with a team of 20 to classify microbiomes in the St. Lawrence River using autoencoders (TensorFlow, Keras) for data normalization, outperforming traditional methods like min-max scaling and z-score normalization in preserving microbial diversity and latent structure

- Generated low-dimensional embeddings of microbiome data, enhancing clustering and classification of microbial communities. Applied unsupervised learning techniques (DBSCAN, t-SNE) to detect rare bacterial strains linked to disease, improving classification accuracy by 17% over conventional approaches

INDUSTRY EXPERIENCE

Wondervogel Pvt Ltd, Kolkata- Data Analyst Intern January 2024 - April 2024

- Led a team of 4 to develop a Graph Neural Network (GNN)-based route recommendation system for Goa, optimizing travel itineraries based on budget and user preferences, increasing customer satisfaction by 35%. Integrated geospatial (OSMNX) and traffic data (NetworkX) with Particle Swarm Optimization (PSO) and Graph Convolutional Networks (GCN using PyTorch-Geometric), achieving 88% accuracy
- Processed 5 years of historical traffic data, using LSTM (TensorFlow/Keras) to model temporal trends, reducing a complex network to 50 key destinations and 82 routes. Optimized travel plans, cutting fuel consumption by 22% and travel time by 18%. Published findings in master's thesis, winning the Best Graduate Researcher Award and contributing to a 12% increase in company profit

Agitechavids Pvt Ltd, Kolkata- Data Analyst Intern June 2019 - July 2019

- Built a grocery recommendation engine using Collaborative Filtering (KNN-based, recosystem R package), Matrix Factorization (SVD via irlba), and XGBoost (xgboost package), achieving 82% accuracy
- Led data preprocessing (Pandas, dplyr), conducted A/B testing (shiny R package), increasing user engagement by 15% and driving a 10% boost in sales through personalized recommendations

TEACHING EXPERIENCE

Teaching Assistant- SURV626: Sampling I September 2025 - November 2025

- Supported instruction in sampling theory, complex survey designs, stratification, clustering, weighting, variance estimation, and design-based inference.
- Graded problem sets and exams, identified conceptual misunderstandings, and provided targeted explanations on PSU selection, FPC, Taylor linearization, and nonresponse adjustment methods.

Teaching Assistant- SURV612: Ethical Consideration for Data Science Research September 2025 - September 2025

- Assisted students in understanding core ethical frameworks in data science, including privacy, bias, informed consent, data governance, and responsible AI use.
- Graded writing assignments and case analyses with detailed feedback, helping students articulate ethical reasoning and apply principles to real-world data scenarios.

RELEVANT PROJECTS & PUBLICATION & CONFERENCE PROCEEDINGS

Multilevel Determinants of Player Market Value in Elite European Football, University of Maryland

- Built a cross-classified multilevel model using 7,000+ player-season observations (2013–2024) across 21 clubs to quantify player-, club-, and time-level effects on market valuation.
- Modeled nonlinear temporal patterns with polynomial terms and random slopes using lme4; conducted LRT-based model comparison, ICC estimation, and full diagnostic evaluation.
- Demonstrated advanced hierarchical modeling, structuring of time-varying vs. time-invariant predictors, and rigorous inference in a complex social-behavioral dataset.

An Empirical Evaluation of GTI Rankings Through Event Data and Media Framing, University of Maryland

- Designed an interpretable machine learning framework to reconstruct the Global Terrorism Index (GTI) by integrating structured GTD event data with media framing extracted from 6M+ news articles using decision trees and random forest regression
- Identified GTI score inflation in countries like India and Syria and uncovered framing biases via feature attribution, residual analysis, and framing-class proportions, leading to a more context-aware and fairer terrorism index. This paper was selected at the **IISA 2025** conference at Lincoln

Social Media Sentiment Dynamics Toward the U.S. in the Context of Zelensky's White House Visit, University of Maryland

- Analyzed over 3M Reddit posts from 10 political and conflict-related subreddits to evaluate public sentiment shifts toward the U.S. before and after Ukrainian President Volodymyr Zelensky's White House visit
- Detected a notable increase in positive sentiment toward the U.S. and the current administration post-visit, with consistently strong pro-U.S. sentiment observed in the r/Ukraine subreddit

Prediction of Public Sentiment using LLM compared to internet data for Electronic Vehicle, University of Maryland

- Conducted sentiment analysis on EVs using the DistilBERT transformer model on internet data (Reddit, X, Instagram, YouTube, Google Search) and traditional sources (The New York Times, Washington Post), comparing results with LLMs like GPT and LLAMA
- Challenged emerging ideas that LLMs accurately predict public sentiment for market research, demonstrating their limitations for EVs; findings revealed BEVs as the most popular and PHEVs as the most negatively viewed. Selected for presentation at the **80th AAPOR Conference**

Topological Data Analysis on the structure of DNA and RNA, Vellore Institute of Technology

- Analyzed DNA and RNA structures using Topological Data Analysis (TDA) techniques such as Mapper (via Kepler Mapper) and Persistent Homology (Ripser), identifying structural patterns and defects with 92% accuracy in classification
- Built topological representations of high-dimensional genomic data to uncover local and global structural insights; awarded Best Paper at the **25th NCSET Conference**

Image Recognition using Factorial Design experiment, Vellore Institute of Technology

- Conducted a full factorial design experiment on CNNs using the CIFAR-10 dataset (60k images), systematically testing 64 hyperparameter combinations (filters:16–128, kernels:3×3/5×5, learning rates:0.001–0.01) to optimize model performance. Achieved 67.72% accuracy (precision: 0.68, F1:0.675) with optimal parameters (3×3 kernel, 128 dense units, 0.001 LR), demonstrating systematic parameter tuning tradeoffs
- Benchmarked 5 neural architectures: DenseNet outperformed others (81.2% accuracy, F1:0.812), while VGGNet (62.1%) and ResNet (38.1%) underperformed. Quantified computational efficiency (CNN: 89s vs. DenseNet:1532s) using TensorFlow/Keras on M1 Pro hardware, revealing architecture-specific accuracy/runtime correlations

Comparative Analysis of PageRank and HITS algorithm using Citation Network Data, Vellore Institute of Technology

- Conducted comparative analysis of PageRank and HITS algorithms on a 28.1k-node citation network using Python (NumPy, SciPy), achieving Spearman rank correlations of 0.968 (PageRank vs. hubs) and 0.891 (authorities vs. random walks) to evaluate algorithmic efficiency
- Identified HITS as superior for citation networks via EDA, revealing gamma-distributed PageRank scores (shape=1.91, rate=5.3M) and hub/authority rank skewness (7.69–13.98), awarded Best Paper at the **24th NCSET Conference**

SKILLS

- **Programming:** Python, R, SQL
- **Machine Learning:** PyTorch, TensorFlow, Scikit-learn
- **NLP:** Hugging Face, spaCy, NLTK
- **Data & Pipelines:** Pandas, dplyr, Playwright
- **Statistical Methods:** causal inference, survey methodology, multi stage regression modelling
- **Tools:** Git/GitHub, AWS/GCP

ADDITIONAL INFORMATION

- **Certification:** Diploma in Data Science- Indian Institute of Technology, Chennai
- **Association Membership:** AAPOR, ASA, RSS
- **University Clubs:** Dean Graduate Student Advisory Council(DGSAC)- Committee Member, UMD GROW Committee, Member, UMD Sports Analytics Club

REFERENCES

Dr. Brian Kim (kimbrian@umd.edu 626-224-8488) – UMD JPSM Professor, Instructor (SURV727, SURV640)

Julia Mendelsohn (juliame@umd.edu 516-457-3193) - UMD INFO Sci Assistant Professor, Research Supervisor

Raphael Nishimura (raphaeln@umich.edu 734-389-4959) - Umich Director of Sampling Operations of the Survey Research Operation