Namit Shrivastava

College Park, MD, USA | (240)-476-8513 | namit507@gmail.com
LinkedIn | GitHub | Website

RESEARCH SUMMARY

Graduate researcher with proven expertise in survey methodology, causal inference, and advanced statistical modeling. Conducted geospatial epidemiological research at the University of Michigan's Institute for Social Research by developing stratified analytical frameworks for large-scale census tract data with complex missing data patterns. Published survey research in peer-reviewed venues (Springer) and presented at AAPOR on transformer-based sentiment analysis methodologies. Academic background combines rigorous statistical training with practical survey research experience, achieving Dean's Fellowship recognition. Research agenda focuses on developing trustworthy data integration frameworks that fuse survey methodology principles with generative AI, deep learning architectures, and privacy-preserving computational methods to advance automated data collection, quality assurance, and responsible measurement at scale.

EDUCATION

University of Maryland, College Park Aug 2024 - May 2026
Master of Science, Survey & Data Science (Data Science Track) College Park, MD, USA

- **GPA: 3.814/4.0** | Dean's Fellowship Award Winner AY 2025-26
- Coursework: Statistical Modeling and Machine Learning (I and II), Fundamentals of Data Collection (I and II), Experimental Design and Causal Inference, Fundamentals of Computational Data Display, Long-Context Language Models, Machine Learning for Social Science, Applied Sampling, Applications of Statistical Modeling, Total Survey Error and Data Quality I, Fundamentals of Inference

Birla Institute of Technology and Science (BITS) Pilani Nov 2020 - Jul 2024
Bachelor of Engineering (Honours), Civil Engineering (Minor: Data Science) Pilani, RJ, India

- **GPA: 3.327/4.0**
- Coursework: Foundations Of Data Science, Object Oriented Programming, Applied Statistical Methods, Artificial Intelligence, Data Mining, Data Visualization, Human Resource Development, Technical Report Writing, Computer Programming, Principles of Management, Electrical Sciences, Machine Learning, Mathematics (I, II and III), Operating Systems, Probability and Statistics, Business Communication, Engineering Graphics, Fluid Mechanics, Construction Plan and Tech, Analysis of Structures

PROFESSIONAL EXPERIENCE

Social Data Science Center, University of Maryland Jan 2026 - May 2026
Graduate Research Assistant College Park, MD, USA

- Engineered Python automation suite interacting with CKAN REST API to **manage lifecycle of 18 datasets** ensuring 100 percent resource accessibility and data integrity
- Architected scalable data taxonomy for university repository transforming flat catalogs into hierarchical thematic groups which **improved search discoverability by 35%**
- Implemented bulk metadata update scripts **reducing manual maintenance overhead by 90%** and

ensuring strict schema compliance across the repository

Institute for Social Research, University of Michigan May 2025 - Dec 2025
Research Assistant Ann Arbor, MI, USA

- Architected and deployed a production-scale geospatial data integration pipeline processing **129,572 U.S. census tracts** across **3 RUCA** (Rural-Urban Commuting Area) strata, achieving **100% broadband data completeness** through multi-source fusion (FCC, ACS, CDC) and reducing baseline missingness by **28.6%** via composite measure imputation.
- Developed and validated a stratified epidemiological modeling framework analyzing COVID-19 incidence patterns across **2,788 census tracts** with **100% social determinant coverage**, uncovering statistically significant non-linear rural inflections ($\mathbf{p < 0.01}$) and heterogeneous broadband effects by urbanicity.
- Designed and implemented a multi-stage data quality assurance protocol integrating Moran's I spatial autocorrelation analysis (identifying **15% spatial clustering violations**), temporal coverage validation (achieving **85% of tracts with $\geq$95% completeness**), and RUCA-stratified missingness diagnostics.

Joint Program in Survey Methodology (JPSM) Feb 2025 - May 2026
Teaching and Graduate Assistant College Park, MD, USA

- Assisted Dr. Jörg Drechsler from Institute for Employment Research in teaching SURV735, effectively guiding **23 students** in understanding data privacy and confidentiality principles.
- Redesigned and standardized Canvas LMS infrastructure for **10+ JPSM instructors** supporting **125+ graduate students** by implementing modular course templates, automated grading workflows, and accessibility-compliant materials. This increased course satisfaction scores by **30%** and reduced instructor setup time by **40%**.

Legistify Services Private Limited Jan 2024 - Jun 2024
Machine Learning Engineer Gurugram, HR, India

- Engineered a scalable logo similarity detection system processing **2.4 million images** for intellectual property infringement analysis. Implemented perceptual hashing (imagehash), feature extraction (SIFT), and approximate nearest neighbor search (Faiss) via FastAPI microservice. Achieved **92% precision** on trademark conflict identification and reduced manual review time by **70%**.
- Deployed production Azure Cognitive Services API leveraging Vision Studio 4.0 for multi-modal document analysis (OCR, object detection, image captioning). Processed **50,000+ legal filings** with **95% extraction accuracy** and enabled bilingual text recognition (English/Hindi) using EasyOCR. This system directly supports **150+ client cases monthly**.
- Developed phonetic trademark similarity algorithm using double metaphone and Jellyfish string matching to compare **50,000 new trademarks** against **300,000 existing database entries**. Achieved **92% accuracy** in conflict detection while reducing computational overhead by **60%** through intelligent prefiltering and optimized SQL queries.
- Integrated and stress-tested Legistrak API ecosystem by executing **150+ REST API calls** via Postman with **95% success rate**. Identified **12 critical bottlenecks** and implemented caching strategies that increased data retrieval efficiency by **30%**. This improved platform responsiveness for **500+ daily active users**.

Accenture Jun 2023 - Aug 2023
Advanced Application Engineering Analyst Bangalore, KA, India

- Monitored enterprise security infrastructure by analyzing threat intelligence feeds from MITRE ATT&CK, AlienVault OTX, and proprietary sources. Achieved **89% accuracy** in threat classification and enabled proactive defense against **25+ emerging vulnerabilities**. This contributed to a **20% reduction** in incident detection latency.
- Conducted penetration testing and vulnerability assessments on **50+ web applications**, networks, and cloud infrastructure using Burp Suite, Nmap, and Metasploit. Identified and documented **120+ security vulnerabilities** (**15 critical, 45 high-severity**) that directly informed remediation roadmaps for Fortune 500 clients.
- Supported incident response operations during **3 security breaches** by performing log analysis with Kusto Query Language (KQL) and Azure Sentinel. Assisted with containment and forensics while contributing to post-incident reports. This reduced mean time to recovery (MTTR) by **80%** through automated runbook implementation.
- Developed Python automation scripts for log parsing, threat indicator extraction, and security metrics dashboarding. Processed **10M+ daily log events** and generated executive-level threat summary reports that enabled data-driven security decision-making for C-suite stakeholders.

Indian Red Cross Society May 2022 - Jul 2022
Web Developer Bangalore, KA, India

- Designed and deployed a Drupal-based content management system for volunteer registry and donor database management. Streamlined volunteer onboarding workflows and reduced manual data entry time by **50%**.
- Localized IRCS main website by implementing multilingual support for Hindi and Kannada regional languages. Conducted comprehensive testing for functionality, performance, and security. Achieved **95% bug-free** user experience across **10,000+ monthly visitors**.

PUBLICATIONS AND PROJECTS

Achieving Sustainability in Supply Chain during Disruption Times: Role of Industry 4.0
Advances in Data-Driven Computing and Intelligent Systems (Springer) Aug 2023 - Feb 2024
BITS Goa, India | [Publication Link]

- Designed and executed a qualitative survey study with **200+ automotive supply chain experts** across India, employing fuzzy-set Analytical Hierarchy Process (FAHP) to rank **5 Industry 4.0 technologies** and **6 green supply chain practices**. Achieved consistency ratios below **0.10 threshold**, validating expert judgment reliability across all matrices.
- Developed a hierarchical Interpretive Structural Modeling (ISM) framework integrating MICMAC analysis to map causal relationships between digital technologies (IoT, Big Data Analytics, Cloud Computing) and sustainable practices. Identified government regulation and top management commitment as key drivers with highest driving power (score of **9 out of 9**).
- Established quantitative linkage between Industry 4.0 adoption and green supply chain performance through three-level hierarchical structure. Demonstrated that IoT, BDA, and cloud computing function as critical linkage factors (**driving power 7, dependence power 7**), enabling effective supplier-customer collaboration and circular economy practices.

Analyzing Public Sentiments of EVs in the Era of Sustainable Transformation
80th AAPOR Annual Conference Nov 2024 - May 2025
St. Louis, MO | [Conference Paper]

- Engineered a multi-source sentiment analysis pipeline processing **1.1M+ social media posts** from

5 Reddit communities (**550K comments**) and **40 New York Times articles**. Implemented DistilBERT transformer model achieving **91.6% classification accuracy** on electric vehicle discourse spanning 2020-2024.

- Discovered systematic bias in large language model sentiment prediction through comparative analysis of **10 Groq LLM variants** (Llama 3.1/3.2 series, Mixtral). Found LLMs exhibited **+0.57 positive sentiment bias** compared to actual Reddit data (**M=-0.18, SD=0.52**), with statistical significance (**F(2,549)=28.43**).
- Identified temporal trends showing **35% increase in negative EV sentiment** on Reddit post-2022 despite rising adoption rates, while NYT coverage remained consistently negative (**M=-0.23**) for core EV terms but positive for Tesla and autonomous keywords. Demonstrated that Llama-3.2-90b-vision-preview best approximated human sentiment patterns.

Player Market Value Analysis in Elite European Football
Graduate Research Project Sep 2025 – Dec 2025
University of Maryland, College Park, MD, USA | [LinkedIn]

- Developed a cross-classified multilevel model analyzing **7,023** player-season observations from **2,159** elite European football players, explaining **78%** player-level variance and showing squad-status playing time drives a **42%** market valuation premium over traditional performance metrics.
- Engineered a cubic polynomial time-series model capturing non-linear market dynamics (**23.7** quadratic contraction; **11** cubic correction), isolating COVID-19's structural impact on the **€38bn** transfer market and quantifying valuation gaps (**65%** goalkeeper discount; **34%** East Asian discount; **21%** Latin American premium; **17%** domestic-citizen discount).

English Premier League (1992–93 to 2021–22): Spending Analysis and 20-Year Forecasting
Course Project – Sports Analytics Jan 2025 – Apr 2025
University of Maryland, College Park, MD, USA | [LinkedIn]

- Automated extraction, cleaning, and integration of **30** EPL seasons (1992–2021) into a unified dataset with **600+** club-season records to support advanced analytics and long-horizon forecasting.
- Built and evaluated Random Forest, ARIMA, Prophet, LSTM, and SVR models, achieving **72.3%** out-of-sample classification accuracy and generating **20-year** forecasts; quantified that higher average annual expenditure corresponds to finishing **2–4** places higher and visualized a **500%** increase in average club expenditure since league inception.

Air Pollution Abatement and Modeling
Undergraduate Thesis Project Aug 2023 – Dec 2023
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Conducted an air quality monitoring study in Pilani using APM 460/APM 550 cyclone-based respirable dust samplers, processing **26** ambient air samples across **766** hours and estimating PM_{10} mean **184.2** $\mu g/m^3$ with peaks **307.6** $\mu g/m^3$; observed exceedance of national standards by **64.5%**.
- Developed ML forecasting models (Linear Regression, Decision Tree, Random Forest) and implemented a Gaussian plume dispersion model achieving $\mathbf{R^2}$**=0.87** with RMSE **12.4** $\mu g/m^3$ for PM_{10} (and $\mathbf{R^2}$**=0.85**, RMSE **8.7** $\mu g/m^3$ for $PM_{2.5}$), evaluating **4** emission-reduction scenarios with up to **24%** projected pollution reduction.

Applications of ML/IoT in Concrete Technology and Structural Health Monitoring
Undergraduate Research Project May 2023 – Dec 2023
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Developed a multilabel CNN using ResNet50/ResNet101 on **42,520** road-surface images (1920×1080) across **5** distress categories, achieving **86.67%** accuracy, **88.19%** F1-score (80/20 split) and reducing

manual road inspection time by an estimated **90%**.
- Expanded training data from **52,009** to **245,735** labeled distress instances via augmentation and deployed a multilabel classifier using binary cross-entropy with sigmoid activation, supporting multi-distress images and enabling scalable road monitoring workflows.

LEXNet: Lightweight Explainable CNNs for Internet Traffic Classification
Undergraduate Research Project Aug 2023 – Dec 2023
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Architected LEXNet for real-time traffic classification across **1.5M** network flows and **200** application classes, matching ResNet-level accuracy while reducing model size by **97%** (to **118,880** parameters) and cutting inference time by **93%** for deployment on resource-constrained routers.
- Engineered LERes and LProto components reducing backbone parameters by up to **19%–36%** and CPU inference time by up to **24%–41%**, while improving accuracy by **4%** and providing faithful prototype-based explanations that outperform ProtoPNet and LSTM/ResNet baselines.

Madhav Gadgil vs. Kasturirangan Reports: Western Ghats Conservation Policy Analysis
Course Project Apr 2023
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Conducted environmental policy analysis comparing two Western Ghats conservation frameworks across **6** states, **44** districts, and **142** taluks; quantified ESA coverage differences (Gadgil: **75%** / **137,000** hectares vs. Kasturirangan: **37%** / **60,000** hectares), a **50.7%** reduction in protected coverage.
- Analyzed governance and zoning approaches (Gadgil ESA-1/2/3 allocation: **60%/15%/25%** vs. Kasturirangan emphasis on **90%** natural landscape protection with **8%** cultural-use allocation), highlighting a shift from bottom-up Gram Sabha governance to centralized working-group execution.

Active Learning for Forest Cover Classification
Course Project – Machine Learning Sep 2022 – Dec 2022
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Implemented an active learning framework on a forest cover dataset of **581,012** observations with **54** features, achieving **85%** accuracy using Entropy Sampling and outperforming random sampling by approximately **15** percentage points for **7** cover types.
- Compared **5** active learning strategies and built a Query-by-Committee ensemble (Random Forest + Decision Trees), demonstrating uncertainty-based sampling reduces labeling needs by **60–70%** while maintaining **82–85%** model accuracy.

Automated Plagiarism Detector for Text and Images
Course Project – Artificial Intelligence Sep 2022 – Dec 2022
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Developed an automated plagiarism detection system implementing the Winnowing algorithm with a dual-search architecture (local database + MSN Live Search), supporting essays, code, and podcast transcripts; achieved **68.1%** accuracy with a LinearSVC classifier and enabled template fingerprint removal to reduce false positives.
- Engineered an image plagiarism module using Average Hash (A-Hash) with 8×8 grayscale hashing and automated Excel reporting with hash-difference visualization, reducing manual plagiarism review time by an estimated **95%**.

Statistical Analysis and Forecasting of Solar Energy
Course Project – Applied Statistical Methods Sep 2022 – Dec 2022
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Engineered a solar energy forecasting system on a **14-year** dataset (2000–2014) with **8** meteorological parameters, identifying correlations exceeding **0.80** for key GHI factors and selecting beta distributions via Kolmogorov–Smirnov testing.
- Built AR/MA/ARMA/ARIMA/SARIMA models for solar radiation prediction, achieving **5.93%** MAPE using SARIMA on **781** weekly observations and demonstrating strong accuracy with lower computational cost than ML baselines.

Voice Gender Recognition using Acoustic Features
Independent Data Science Project Jul 2022
Birla Institute of Technology and Science (BITS) Pilani, Pilani, Rajasthan, India | [LinkedIn]

- Developed a gender recognition model on **3,168** voice samples with **20** acoustic features, benchmarking **5** ML algorithms and achieving **98.3%** classification accuracy using Random Forest (50 estimators).
- Tuned a Random Forest classifier (entropy criterion; max_depth=6; n_estimators=50) achieving **98.6%** training and **98.3%** testing accuracy on a balanced dataset (**1,584** male / **1,584** female), with only **11** misclassifications.

TECHNICAL SKILLS

Programming: Python, R, Java, C++, JavaScript, TypeScript, HTML/CSS, Bash/Shell
Data & Databases: SQL (MySQL, PostgreSQL), NoSQL (MongoDB, Cassandra), Snowflake, Neo4j, Pinecone, Apache Spark, Kafka
AI/ML: PyTorch, TensorFlow, Keras, Hugging Face, LangChain, LlamaIndex, LangGraph, PySpark, NLTK, Spacy, SkLearn
DevOps: Git, Docker, Kubernetes, Jenkins CI/CD, REST APIs, Django, Flask, Angular, React.js, Terraform, Ansible
Cloud: AWS (EKS, S3, EC2, Lambda, Sagemaker), Azure, GCP, Tableau, Power BI, MLFlow, Qualtrics, STATA
Core: LLMs, Generative AI, RLHF/DPO, Deep Learning, NLP, Computer Vision, Survey Methodology, Causal Inference, MLOps

LEADERSHIP & SERVICE

Terrapin Leadership Institute, University of Maryland Aug 2024 - May 2025
Member College Park, MD

- Achieved **100% participation** in workshops on leadership, ethics, inclusion, and resilience.
- Enhanced group dialogue quality by contributing reflective insights in all sessions, fostering more inclusive and engaging discussions among peers.

National Service Scheme (NSS), BITS Pilani Mar 2022 - Dec 2023
Executive Committee & Blood Donation Camp Core Team Member Pilani, RJ, India

- Planned **10+ activities** to improve English skills and confidence in villages around campus.
- Organized blood donation camp coordinating **60+ volunteers**, managing data for **1,000+ donors**, achieving **844 successful donations** within 2 days.

Peer Mentorship Program (PMP), BITS Pilani Aug 2021 - Dec 2023
Mentor Pilani, RJ, India

- Assisted Juniors with advice and materials to ensure they have a smooth transition into college life and succeed academically.

CERTIFICATIONS & ACHIEVEMENTS

Certifications: Microsoft Certified: Azure AI Fundamentals | API-based Programming (Postman) | AI For Everyone (Coursera) | Deep Learning & Machine Learning (Smartknower, Internshala)

Achievements: **1st rank** in Human Resource Development (**180 students**, BITS Pilani) | **Top 10** in Water & Wastewater Treatment (**98 students**, BITS Pilani) | JPSM Dean's Fellowship Award (AY 2025-26, UMD)