MANAN SAXENA

+1(814) 769-0852 ◇ State College, PA (Willing to relocate)
manansaxena05@gmail.com ◇ linkedin.com/in/manansaxena05 ◇ manansaxena.github.io

EDUCATION

Pennsylvania State University, University Park, PA — January 2025 - May 2027
Doctor of Philosophy, Informatics (Data Science Concentration)

Pennsylvania State University, University Park, PA — August 2022 - May 2024
Master of Science, Informatics (Data Science Concentration), GPA: 3.92/4.0

Delhi College of Engineering, India — August 2016 - June 2020
Bachelors of Technology, Software Engineering, GPA: 3.7/4.0

AREAS OF EXPERTISE

- Statistical Modeling
- Machine Learning
- Time Series Forecasting
- Computer Vision
- Deep Learning
- Cloud Computing
- Web Development
- Generative AI

EXPERIENCE

Research Associate, *Pennsylvania State University, University Park, PA* — May 2023 - Present
Supervisor: Prof. Justin Silverman

- Developed a scalable Bayesian inference algorithm for multivariate count time series data, applied to understanding trends and patterns in microbial systems.
- Calculated closed-form gradients for posterior estimation, achieving **20-30x** faster optimization than automatic differentiation in Stan, and generated 95% credible intervals using Multinomial Dirichlet Bootstrap with almost **0** deviation from the true posterior.
- Created an R package called **Fenrir** with base code in C++ utilizing Eigen and Boost libraries for optimized performance. **[Link]**
- Engineered a codebase employing shell scripts to run automated jobs with minimal user input on Penn State's High-Performance Compute (HPC). **[Link]**

Software Engineer, *Tummee.com*, *Remote* — September 2021 - June 2022

- Managed end-to-end software development of a feature addition to the core sequence builder functionality of the platform
  - Ensured **0 fault** live deployments and optimal cross-platform performance.
  - Secured over **800 users** within a **month** of release.
- Led a cross-functional team to revamp the customer issue submission portal, leveraging customer insights to improve user experience, resulting in a **20%** reduction in issue resolution time.
- Developed REST-APIs in Python's webapp2 framework integrated with GCP. Built UIs using Bootstrap and handled dynamic behavior with JavaScript and jQuery.
- Automated the transformation of unstructured data to the structured database using Google Sheets API, reducing manual data entry time by **50%**.

Machine Learning Engineer, *Trinity College Dublin, Leinster, Ireland* — June 2019 - July 2021
Supervisors: Prof. Ciaran Simms, Prof. Aljosa Smolic, Richard Blythman

- Developed an automated end-to-end pipeline using fine-tuning of deep learning models for predictive analytics in sports injury prevention. Used 3D pose estimation, object tracking, and instance image segmentation models.
- Tested proof-of-concept level model on novel rugby tackle data set, comparing to industry benchmark motion capture systems (VICON) with reasonable performance and at a fraction of the cost.
- Built an automated pipeline for camera calibration and face blurring to acquire rugby tackle datasets. **[Link]** **[Link]**
- Collaborated with coaches and physiotherapists to translate domain knowledge into decisions for prototype development, resulting in more effective injury prevention strategies.

SKILLS

Programming Languages	C++, Python, R, SQL, JavaScript, SAS, Stan
Development Tools	Google Cloud Platform, Databricks, jQuery, Docker, GitHub, Linux, Jenkins
Data Science and AI Tools	PyTorch, TensorFlow, Keras, LangChain, MLflow, Streamlit, Statsmodel, Prophet, Scikit-Learn, OpenCV, Tidyverse, Tableau

PUBLICATIONS

- **Saxena, M.**, Chen, T., & Silverman, J. D. (2024). Scalable inference for Bayesian multinomial logistic-normal dynamic linear models. arXiv preprint, arxiv.org/abs/2410.05548.
- Blythman, R., **Saxena, M.**, Tierney, G. J., Richter, C., Smolic, A., & Simms, C. (2022). Assessment of deep learning pose estimates for sports collision tracking. Journal of sports sciences, 40(17), 1885-1900. doi.org/10.1080/02640414.2022.2117474
- Dhiman, C., **Saxena, M.**, & Vishwakarma, D. K. (2019, September). Skeleton-based view invariant deep features for human activity recognition. In 2019 IEEE fifth international conference on multimedia big data (BigMM) (pp. 225-230). IEEE.
- Garg, A.*, Aggarwal, K.*, **Saxena, M.***, & Bhat, A. (2021). Classifying medical histology images using computationally efficient CNNs through distilling knowledge. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3 (pp. 713-721). Springer Singapore.

TALKS

Scalable Inference for Bayesian Multinomial Logistic-Normal Dynamic Linear Models October 2024

- Bioinformatics Method Developers Community Day, Center for Computational Biology and Bioinformatics, Penn State

PROJECTS

Baseball: Modeling Batter's Swing Probability June 2024

- Developed a predictive model for batter's swing probability for different types of pitches using a Random Forest classifier and Grid Search Cross Validation, achieving an accuracy score of **89**%.
- Introduced the Swing Efficiency Index (SEI) metric, which measures the ratio of actual swing percentages to the adjusted swing probability based on pitch height, providing insights into the batter's mentality and execution. Additionally, analyzed middle-middle pitches and presented findings in a format understandable by coaches. **[Link]**

Forecasting SARS-CoV-2 Concentrations in Wastewater May 2024

- Created and deployed a time-series forecasting model for SARS-CoV-2 Concentrations using Prophet integrated with MLflow for tracking, improving **RMSE** from **140** in the trend model to **125**. Built containerization and CI/CD pipeline development using Docker and Jenkins, facilitating deployments to GitHub Container Registry. **[Link]**

Music Lyrics Analysis and Q&A System February 2024

- Deployed an interactive system on Streamlit for analyzing and responding to queries about music lyrics. Utilized LangChain framework, combined with ChatGPT (Large Language Model) API for natural language processing (NLP) and YouTube API for lyric extraction from music videos. **[Link]**

Classification of Breast Cancer Histology Images through Distilling Knowledge August 2019 - June 2020

- Implemented a light CNN model for high-resolution breast cancer histology image classification, utilizing knowledge distillation techniques and attention maps. Leveraged ResNet 50 as a teacher model to improve the performance of a lighter ResNet 8 model, boosting its accuracy from **75%** to **80%**.

Skeleton-Based View Invariant Deep Features for Human Activity Recognition December 2018 - May 2019

- Introduced novel view-invariant skeletal features to describe spatial-temporal characteristics of human motion. Achieved a **2%** accuracy improvement over existing state-of-the-art models on the NUCLA dataset through the application of transfer learning and dynamic image techniques.

CERTIFICATIONS AND PROFESSIONAL AFFILIATIONS

2024 Databricks Machine Learning Professional (In progress)
2024 Databricks Fundamentals **[Link]**
2024 Member of American Statistical Association (ASA)

HONORS AND AWARDS

2016 All India Rank **2,661** out of **1.4 million** candidates in Joint Entrance Examination Mains (JEE Mains).
2012 National Talent Search Examination scholar, ranked among the top **1,000** out of **1.2 million** candidates, was awarded a scholarship until the completion of undergraduate studies.