

MANAN SAXENA

+1(814) 769-0852 ◊ State College, PA (Willing to relocate)

manansaxena05@gmail.com ◊ [linkedin.com/in/manansaxena05](https://www.linkedin.com/in/manansaxena05) ◊ manansaxena.github.io

EDUCATION

Pennsylvania State University, University Park, PA Aug. 2022 - May 2024
Master of Science, Informatics (Data Science Concentration), GPA: 3.92/4.0
Delhi Technological University (Formerly DCE), India Aug. 2016 - June 2020
Bachelors of Technology, Software Engineering, GPA: 3.7/4.0

EXPERIENCE

Statistical Researcher, *Pennsylvania State University, University Park, PA* May 2023 - Present

Supervisor: Prof. Justin Silverman

Key Skills: Statistical Modeling, Time-series Analysis, Multivariate Analysis, Forecasting, Machine Learning

- Developed scalable Bayesian inference model and performed hypothesis testing for multivariate longitudinal data with application to understanding patterns of microbiota variation in the human gut.
- Calculated closed-form gradients for parameter estimation resulting in **1.5x** faster performance than automatic differentiation in Stan and generated 95% credible intervals of the estimated parameter using Bayesian Bootstrap.
- Created a C++ header library with an R interface for parallelized sampling that is **32x** faster than base R.
- Thesis defended, preparing for submission to the *39th AAAI Conference on Artificial Intelligence*.

Software Development Engineer, *Tummee.com*

Sept. 2021 - June 2022

Keywords: Fullstack Web Development, Project Management, Problem-solving and Analytics

- Managed end-to-end software development of a feature addition to the core sequence builder functionality of the platform
 - Ensured **zero-fault** live deployments and optimal cross-platform performance.
 - More than **800** users within a **month** of release
- Led a cross-functional team to revamp the customer issue submission portal, leveraging valuable customer insights.
- Developed the backend in Python's webapp2 framework integrated with Google Cloud Platform and used SQL for data querying and manipulation.
- Wrote scripts to automate the transfer of unstructured data to the structured database using Google Sheets API.

Data Scientist, *Trinity College Dublin, Ireland*

June 2019 - July 2021

Supervisors: Prof. Ciaran Simms, Prof. Aljosa Smolic, Richard Blythman

Keywords: Deep Learning, Computer Vision, Machine Learning, Biomechanical Analysis

- Developed an automated end-to-end pipeline leveraging cutting-edge deep learning models for predictive analytics in sports injury prevention, enhancing player safety and performance through data-driven decisions.
- 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 cost.
- Built an automated pipeline for camera calibration and face blurring to acquire rugby tackle datasets using Python.
- Collaborated with coaches and physiotherapists to translate domain knowledge into decisions during prototype development.
- **Published in** *Journal of Sports Science*, 2022. [[Paper](#)]

PROJECTS

Classification of Breast Cancer Histology Images through Distilling Knowledge

- 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%**.
- **Published in** *Lecture Notes in Networks and Systems, vol 164. Springer, Singapore.* [[Paper](#)]

Forecasting SARS-CoV-2 Concentrations in Wastewater

- Created and deployed a time-series forecasting model for SARS-CoV-2 Concentrations using Prophet integrated with MLflow for tracking. 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

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

SKILLS

Programming Languages	C++, Python, R, SQL, JavaScript, Stan
Web Development Tools	Flexbox, Bootstrap, webapp2, Google App Engine, Google Datastore, jQuery, Docker, Git, Linux, Jenkins
Data Science and AI Tools	PyTorch, Keras, LangChain, MLflow, Streamlit, Statsmodel, Prophet, Pandas, Scikit-learn, Scipy, Matplotlib, Seaborn, OpenCV, Rcpp, Tidyverse, Tableau

AWARDS AND AFFILIATIONS

- 2024** Member of American Statistical Association (ASA)
- 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.