Sameer Shankar

Data Analyst @ BC Cancer | *Vancouver, BC, Canada*
sameer.shankar01@gmail.com +1 (604)-440-3522 LinkedIn GitHub

Education

Bachelor of Science (BSc.), Combined Major in Statistics and Economics Sep 2019 – May 2024
The University of British Columbia (UBC), Vancouver, Cumulative GPA: **A**, *Dean's Honour List*

Research Interests

- Natural Language Processing
- Healthcare applications of Machine Learning/Deep Learning/Reinforcement Learning
- Computer Vision
- Statistical Modelling
- Bayesian Statistics
- Sports Analytics

Research Experience

Data Scientist - Bioinformatics – *R and Python, Data Wrangling and Visualization* Aug 2022 – Aug 2023
BC Cancer, Vancouver

- Co-author and statistician for uncovering therapeutic targets and prognostic markers in p53abn endometrial cancer under Dr. David Huntsman, applying sWGS and targeted sequencing to advance treatment strategies
- Rigorously developed, tested, and evaluated a complex algorithm to identify the clusters in bivariate plots, achieving 72% accuracy under the guidance of Dr. Ryan Brinkman, supporting the discovery of new drug therapies; this work later became a presentation

Data Scientist - NLP Researcher – *Python, Natural Language Processing* Sep 2023 – Dec 2023
UBC Data Science Institute, Vancouver

- Automated classification of histology, neoadjuvant therapy, and behaviour with over 95% accuracy using NLP models, enhancing efficiency in medical data analysis under Professor Raymond Ng; the final model was deployed in production

Publications

Journal Articles

- Jamieson, A., Sobral de Barros, J., Cochrane, D. R., Douglas, J. M., **Shankar, S.**, Lynch, B. J., Leung, S., Martin, S., Senz, J., Lum, A., Drew, Y., Gilks, C. B., Huntsman, D. G., & McAlpine, J. N. (2024). Targeted and Shallow Whole-Genome Sequencing Identifies Therapeutic Opportunities in p53abn Endometrial Cancers. **Clinical cancer research**: *an official journal of the American Association for Cancer Research, 30*(11), 2461–2474. https://doi.org/10.1158/1078-0432.CCR-23-3689
- (Awaiting review) **Shankar, S.** (2024). An Econometric Analysis of Dietary Diversity in Bangladesh's Farm Households: An in-depth Panel Study. **IONA Journal of Economics**.

Notable Courses/Teaching Experience

Senior Tutor Sep 2022 – Aug 2024
Alma Mater Society of the University of British Columbia, Vancouver

- Completed the following courses with an A and tutored over 100 students in Mathematics, Statistics, Computer Science, and Economics, significantly improving their grades:
  - MATH **200** (Multivariable Calculus), 220, **221** (Matrix Algebra)
  - STAT 300, 302, **305** (Bayesian Statistical Inference), 306, 404, **443** (Time Series Analysis)
  - CPSC 221, 304, **340** (Machine Learning and Data Mining)
  - ECON 323, **407** (Computational Macroeconomics), **425** (Econometrics)

Extracurricular Research Experience

Data Analyst – *Computer Vision, using PyTorch and OpenCV | Repo* Jul 2022 – Aug 2024
UBC Bionics, Vancouver

- Implemented AlexNet Convolution Neural Network to map objects to grips for prosthetics and pruned YOLOv8 with minimal performance loss to fit on a Raspberry Pi, aiming to assist amputees

Notable Course Projects

ECON 490 – Seminar in Applied Economics
An Econometric Analysis of Dietary Diversity in Bangladesh's Farm Households – *R* Jan 2024 – Apr 2024

- Conducted an econometric panel analysis of dietary diversity in Bangladesh's farm households, utilizing fixed and random effects models, finding that a 10% increase in farm income proportion leads to a statistically significant 3% increase in dietary diversity score
- Gained expertise in developing a research paper, handling complex panel datasets and applying advanced econometric techniques, contributing to actionable insights for policymakers

STAT 447B – Statistical Computing
Machine Learning models to predict building energy consumption – *Python | Repo* Jan 2022 – Apr 2022

- Trained ML models, including Random Forest, SVM, ElasticNet, XGBoost, and LGBM Regressor, achieving a competitive RMSE of 18.501 (Top 30 in Kaggle), authoring a research-style group report
- Incorporated hyperparameter tuning, PCA, and KNN imputation techniques, to identify buildings requiring retrofitting, thereby enhancing energy efficiency and reducing maintenance costs

Technical Skills

Languages: Python (NumPy, SciPy, Pandas, Seaborn, Matplotlib, scikit-learn, TensorFlow, PyTorch, Non-conformist), Java, R (Tidyverse, Bioconductor), Julia, MATLAB, Tableau, SQL, Apache Spark, Django, Heroku
Competencies: Clustering, Linear Regression, Regularization, Feature Engineering, Kernel Trick, Stochastic Gradient Descent, Gradient Boosting, Maximum Likelihood Estimation, Principal Component Analysis, Neural Networks, Backpropagation, LLM Transformers, Relational Databases, Algorithms and Data Structures, Quantitative Modelling

Service

Senior Data Analyst – *Python, R, and Excel, Data Cleaning and Visual Analytics* Jul 2022 – Apr 2024
Vancouver School of Economics Undergraduate Society (VSEUS), Vancouver

- Created the VSEUS budget, analyzed events, and recommended strategies to boost turnout to 90%+

Statistics Undergraduate Liaison Officer Sep 2022 – Jun 2023
Department of Statistics, Vancouver

- Liaised between the department and undergraduates, advising juniors on careers in Statistics

References

Professor Dr. Raymond Ng
Dept. of Computer Science, University of British Columbia
rng@cs.ubc.ca
604-822-2394

Professor Dr. David Huntsman
Dept. of Pathology and Laboratory Medicine and Obstetrics and Gynaecology, University of British Columbia
dhuntsma@bccancer.bc.ca
604-675-8211

Professor Dr. Hiroyuki Kasahara
Vancouver School of Economics, University of British Columbia
Hiroyuki.Kasahara@ubc.ca
604-822-4814