Sameer Shankar

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*

Work Experience

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

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

- Rigorously developed, tested, and evaluated a complex algorithm to identify the clusters in bivariate plots, achieving 72% accuracy, supporting the discovery of new drug therapies
- Co-author and statistician for uncovering therapeutic targets and prognostic markers in p53abn endometrial cancer, applying sWGS and targeted sequencing to advance treatment strategies

Technical Projects

Machine Learning Models to Predict Building Energy Consumption – *Python* 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

Bank Account Fraud Analysis – *Python, Cleansed and Extracted Features* Nov 2022 – Dec 2022

- Developed and implemented a fraud detection system with Logistic Regression (L1 penalty) for feature selection and Isolation Forest for anomaly detection, achieving 94.23% accuracy
- Optimized feature subsets via forward selection and validated using k-fold cross-validation
- Evaluated performance with precision and recall metrics from the Confusion Matrix, effectively mitigating fraud to protect businesses and consumers from financial losses

Regression Analysis of Crime in Denver Neighbourhoods – *R, Statistical Inference* Nov 2021 – Dec 2021

- Applied statistical techniques, including residual plots, to transform variables and optimize AIC/BIC, obtaining lower residual standard error (0.319)
- Developed models based on Adjusted-R^2 (0.728) and Mallow's C_p, validating findings using quantile-quantile plots to ensure robustness and support crime reduction policies

Extracurricular Experience

Data Analyst – *Computer Vision, using PyTorch and OpenCV* 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

Technical Skills

Languages: Python (NumPy, SciPy, Pandas, Seaborn, Matplotlib, scikit-learn, TensorFlow, PyTorch, Non-conformist), Java, R (Tidyverse, Bioconductor), Julia, MATLAB, SQL, Apache Spark, Django, JSON, Heroku
Competencies: Linear Regression, PCA, Neural Networks, LLM Transformers, Bayesian Statistical Inference
Relevant courses: Machine Learning, Relational Databases, Algorithms and Data Structures, Maximum Likelihood Estimation, Econometrics, Quantitative Modelling with Data Science, Computational Economics