Sameer Shankar

BSc. Combined Major in Statistics and Economics | GPA: 83.8%
sameer.shankar01@gmail.com +1 (604)-440-3522 LinkedIn GitHub

Skills

Languages: Python (NumPy, SciPy, Pandas, Seaborn, Matplotlib, scikit-learn, TensorFlow, PyTorch, Non-conformist), Java, R (Shiny, Bioconductor), MATLAB, SQL, Apache Spark, Django, JSON, Heroku
Web: Canva, Adobe Premiere Pro, Adobe Lightroom, Adobe After Effects
Competencies: Data Cleaning, Data Visualization, Random Forest, SVM, Ensemble Models, Gradient Boosting, Time Series Analysis, Anomaly Detection, Bayesian Statistics, Neural Networks

Work Experience

Bioinformatician Aug 2022 – Aug 2023
BC Cancer, Vancouver

- Rigorously developing, testing, and evaluating a complex algorithm used to identify the number of clusters in a bivariate plot, obtaining an accuracy of 72%
- Automating the Genome Analysis Toolkit pipeline for 202 samples using bash scripting and SLURM
- Debugging, documenting, and writing code for 3 functions for the Utanos package development

Technical Projects

Machine Learning models to predict building energy consumption Jan – Apr 2022

- Trained ML models including Support Vector Machines, ElasticNet, XGBoost, LGBM Regressor, and Gradient Boost to get a competitive RMSE of 18.501 (Top 30 in Kaggle Leaderboards), incorporating Hyperparameter Tuning, PCA, KNN Imputation and Random Forest techniques

Bank Account Fraud (BAF) Analysis Nov – Dec 2022

- Found the best subset of features to predict BAF occurrence, based on Logistic Regression with L1 penalty and Isolation Forest; obtained an accuracy of 94.23%
- Implemented ML techniques such as forward selection and k-fold cross-validation, while using evaluation metrics such as precision and recall (based on Confusion Matrix).

Regression analysis of crime in Denver neighbourhoods Nov – Dec 2021

- Used statistical techniques such as residual plots to transform variables to obtain a lower RMSE (97.142), residual standard error (0.319) and AIC/BIC
- Developed model based on Adjusted-R^2 (0.7282) and Mallow's C_p as metrics, and validated the model using a quantile-quantile plot to make findings more robust

Education

Bachelor of Science: Combined Major in Statistics and Economics (4th year), Expected graduation: 2024, The University of British Columbia

- **Computer science courses:** Systematic Program Design, Software Construction
- **Statistics courses:** Intermediate Statistics for Applications, Probability Theory and Combinatorics, Maximum Likelihood, Regression Analysis, Sample Surveys, Design and Analysis of Experiments, Time Series and Forecasting, ML Techniques with Applications
- **Mathematics courses:** Calculus III, Linear Algebra, Applied Linear Algebra
- **Economics courses:** Introduction to Econometrics, Computational Methods in Macroeconomics