Steve Bicko Cygu

Research Goals

I am an applied statistician with strong background in mathematics, computing and data science. My focus is on application of machine learning tools in trying to understand public health challenges. Specifically, using machine learning methods to predict survival of cancer patients.

Profile Summary

A Data Scientist with both strong mathematical, computational skills and over 5 years' in active research, predictive modeling, data processing and mining, algorithms implementation and machine learning to solve public health and data handling challenges. Greatly involved in R open source community and passionate about development of R packages. I desire to work in a highly competitive and ambitious organisation to bring into good use my strong academic qualifications in Applied Statistics, Computing and Predictive Modelling altogether to objectively organize, manage, analyse and make inferences from complex data to reach meaningful conclusions that bridge gaps between people and the data.

Selected Achievements

- R package (pcoxtime) for penalized survival analysis (available on CRAN)
- Automated data quality and reporting R scripts
- R package for the clinicians to estimate age-related in health South African children (available on github)
- R scripts for scrapping web pages and Twitter to extract data and store them in usable formats (web dashboards) ready for analysis
- Advanced bash script to automatically download data from ODK Aggregate server to avoid redundant signing-in through the web-browser

Skills

- Data science: R, C++, Python
- Data analysis: R, SPSS, Stata
- Data visualization: R shiny and dashboards, Microsoft Excel, PowerPoint
- Questionnaire scripting: ODK, Kobo and Google form
- Data cleaning, modeling, and mining
- Machine learning

Education

2018 – Date	**PhD in Computational Science and Engineering, McMaster University, Canada.**
2016 – 2017	MSc in Mathematics, Stellenbosch University, South Africa.
2014 – 2015	MSc in Mathematical Sciences, University of Cape Town, South Africa.
2009 – 2013	BSc in Applied Statistics with Computing, Maasai Mara University, Kenya.
2004 – 2007	Kenya Certificate of Secondary Education, Waondo Secondary School, Kenya.

Research Experience

2018 – Date	**Doctoral Candidate, McMaster University, Canada.**

- Using Machine Learning Methods to Predict the Risk of Death of Gastrointestinal Cancer Patients
- Modeling approaches for multivariate binary response. A case study to investigate the contribution of demographic, social and economic factors to improved water, sanitation and hygiene among the urban poor

2016 – 2017	**Graduate Student, Stellenbosch University, South Africa.**

- Immune biomarker reference range estimation for healthy paediatric patients in South Africa
- Determining paediatric Immune Biomarker Reference Ranges using a model-based, age-continuous estimation method

Publications

In review	**Cygu, Steve**, and Benjamin M. Bolker. *"pcoxtime: Penalized Cox Proportional Hazard Model for Time-dependent Covariates." arXiv preprint arXiv:2102.02297 (2021)*
Manuscript	**Steve Cygu**, Helen Payne, Denise Lawrie, Debbie Glencross, Martin Nieuwoudt. *Determining paediatric Immune Biomarker Reference Ranges using a model-based, age-continuous estimation method*

Teaching Experience

2018 – Date	**McMaster University, Canada.**

Teaching assistant for

- Introduction to mathematical and scientific programming
- Population ecology

Professional Experience

May, 2021 – Date	**African Population and Health Research Center (APHRC), Kenya.** Consultant (Statistician I) • Data mining and web scrapping to create food system database • Data cleaning and management • Sampling • Automated data quality checks • Insightful analysis of data collected
May – Aug., 2018	**African Population and Health Research Center (APHRC), Kenya.** Data Analyst • Developed data visualization tools in R • Developed tools to monitor and clean in coming data in realtime • Insightful analysis of data collected • Questionnaire scripting and development of online based data collection tools • Data cleaning and management
Mar, 2017 – May., 2018	**Dalberg Research, Kenya.** Data Processing Manager • Developed automated data quality check scripts • Coordination and management of data management team • Provided technical assistance to clients • Data analysis and management • Developed data management SOPs
2016 – 2017	**South African Centre of Excellence in Epidemiology Modelling and Analysis, South Africa.** Research Student • Developed statistical and mathematical models for the analysis of epidemiological data • Developed R software package for the estimation of reference ranges
Jan., 2014 – Sept., 2014	**Infotrak Research and Consulting, Kenya.** Data Processing Manager • Designed survey tools • Supervised data collection and entry process • Managed databases and provided realtime update on survey progress • Conducted data cleaning, processing and analysis • provided technical assistance to business development team

Academic Administrative Experience

2012 – 2013	Student Representative (Academic Director), Maasai Mara University.
2011	Treasurer Information Club, Maasai Mara University.
2010	Chairman Mathematics club, Maasai Mara University.

Awards

2016	South African Centre of Excellence Epidemiological Modeling and Analysis Scholarship
2014	frican Institute for Mathematical Sciences (Next Einstein Initiative) Scholarship

Conferences Attended

2019	Compute Ontario HPC Summer School, Hamilton, Canada
2016	Bayesian Analysis of Longitudinal Studies, Stellenbosch, South Africa
	Quantitative Bias Analysis with Epidemiological Data, Stellenbosch, South Africa
	Mathematical Modeling for Infectious Diseases, Cape Town, South Africa
	Clinic on Meaningful Modeling of Epidemiological Data (MMED)

Memberships

2018 – Date	Kenya Statistical Society
2016 – Date	South African Statistical Association (SASA)

Languages

Swahili	Fluent
English	Fluent