

Moses Mburu

Location: Kilifi, Kenya

LinkedIn: <https://www.linkedin.com/in/mmburu/>

Blog: <https://m-mburu.github.io/>

Data Analyst

Data Analyst, with substantial experience in data management, statistical analysis and machine learning. I have worked in healthcare with a bias towards malnutrition, financial diaries research in low income areas and agricultural research. I have gained in-depth knowledge of using statistical/programming languages such as R, Python and SQL for analysis. I enjoy using data to solve problems and I'm always researching on new methods/tools/skills to help me become efficient in my work.

EMPLOYMENT HISTORY

Data Manager, KEMRI WELLCOME TRUST May 2019 – present:

- Role covers a mixture of data management, data analysis and machine learning. In my current role I work under the CHAIN Network where we are focused on optimizing the management and care of the sick and undernourished children in resource-limited settings to improve survival, growth and development. CHAIN study has 9 sites in Africa and Asia.

Data Management:

- I was in a team that developed dashboards for data management using R programming and shiny server. This was achieved by creating automated scripts that were connected to MySQL database and Redcap to check spurious values and for reporting. This ensured that we had an automated way to flag outliers, validation and automation of report generation. This led to a very low data entry error rate of less than 2%. It also helped in that we had a continuous process for data cleaning which made it easier to catch most of the errors early.

Management/Mentorship:

- Taught data clerks and lab technicians on using dashboards to double check entry outlier values or generate reports. I also encouraged them to use project management software to raise tasks to the data team which led us to know early when problems occurred.

Data Visualization:

- Create maps to display malnutrition patterns/prevalence in study countries to see the most vulnerable communities, using R packages such as sf, raster, ggplot2, leaflet and tmap.
- Used plotly and ggplot2, shiny to visualize various study variables patterns and trends. This helped us to dedicate more time to discuss results, research more and try different models for fitting data.

Data Mining:

- Extraction of data sets from various health demographic data bases such as demographic and health surveys database, world population database, National Oceanic and Atmospheric Administration to derive various processed variables. This was useful in finding patterns between weather and hospital visits, how diseases are distributed spatially, how spatial distribution of diseases has changed over years.

Machine Learning:

- Working closely with clinicians to create models that predict high risk children using clinical features, water sanitation and hygiene variables. Clustering methods such as k means clustering was used to group patients that are similar given clinical measurements of features such as complete blood count, blood biochemistry and symptoms. This is important as it helped identify features of high-risk groups.

Spatial Statistics:

- Working closely with the population health unit to model travel times and travel distances to hospitals for study participants.

Data Analyst/Statistician, Low Income Financial Transformation June 2016 – May 2019:

- Low-Income Financial Transformation (L-IFT) is a for-profit social business. The company specializes in a diaries research methodology that can be applied to a range of purposes, such as impact measurement, product development, customer satisfaction gauging, and programme design. L-IFT is primarily focused on financial inclusion, digital finance and strengthening the financial sector of the countries where it works. L-IFT also has expertise and is building data in the fields of energy, livelihoods, youth, entrepreneurship, SME development and the intersection of health and financial management.
- At L-IFT I received a lot of experience in finances of low-income areas. How fluctuating the finances are, savings patterns and how calamities affect they lives.

Data Management:

- This is important step in research as it ensured that the data used in analysis is of the highest quality. This is was achieved by creating data collection databases and reviewing each question with the data collection teams. Monitoring of surveys, flagging inconsistencies in the data entry and raising any queries with the researchers in the field.

Data Mining/Statistical Computation:

- Our surveys were done after every two weeks with a follow up period of more than 6 months. This meant that the data sets produced were large, complex and messy. I Introduced R programming in our company to solve this. This helped with the analysis as we could do more complex analysis for instance digging into passed data and making comparisons to see the trends. R being an open source meant that we had access to actively developed packages for all kinds of analysis. It also meant we had power to do tasks such as visualizing all data variables using frameworks such as shiny.

Data Analysis/Machine Learning:

- This was my day to day job. I worked with the IT department to have shiny server installed in our servers which meant that we could produce come up with dashboards. We were able to stream twitter data in real time. This enabled us to advise on the trending topics that our company could take advantage of in terms of advertising.
- Analysing twitter data to see the impact of influencers by using metrics such as follower's quotient, interactions to user tweets for instance the summary statistics of number of retweets or favourites a user has received over a duration of time.
- Used text analysis methods such as sentiment analysis, topic modelling on open ended questions from surveys.
- I researched extensively on the best ways to visualize and analyse different kind of data sets. This helped us use methods such clustering to identify groups of clients for microfinance institutions in low income settings.

QUALIFICATIONS

Bachelor of Science in Statistics, University of Nairobi, 2012 – 2016:

- Area of study included Exploratory data analysis, Statistical modelling, computation and data analysis, analysis and design, Time series analysis

Spatial Data: Data Camp

- Area of study included visualizing spatial data using ggplot2, tmap and leaflet, spatial analysis with sf and raster, interactive maps with leaflet and spatial statistics, such as point pattern analysis, areal statistics and geo-statistics

Machine Learning: Coursera

- Area of study included supervised learning methods such as regression methods such linear regression, tree-based methods, classification, logistic regression support vector machines tree-based methods neural networks, dimension reduction with pca, clustering methods such k means clustering.

Publications

- Peter Gachoki, Moses Mburu, and Moses Muraya, “Predictive Modelling of Benign and Malignant Tumors Using Binary Logistic, Support Vector Machine and Extreme Gradient Boosting Models.” American Journal of Applied Mathematics and Statistics, vol. 7, no. 6 (2019): 196-204. doi: 10.12691/ajams-7-6-2