PHILIP MUNYUA
Sayreville, NJ 08872
(347)944-9138
mutumamunyua@gmail.com
LinkedIn Profile: https://www.linkedin.com/in/philmunyua

SUMMARY

A PhD trained unicorn Data Scientist that is responsible for data mining, modeling and analytics and providing innovative business intelligence (both short term and long term strategic business insights) to Executive Leadership and Business Unit Managers. My career is founded on executing leadership in data science practice spanning across functional business and finance units; investments, insurance, consumer welfare, marketing, product, sales, and IT (engineering). My strengths, according to independently evaluated 'CliftonStrength' test by Gallup, include; Individualization, Learner, Restorative, Achiever and Includer where I work independently or build teams and draw up the best in each team player for efficient delivery of business insights/innovations that lead to transformative growth across corporations (efficiency-increased revenues and cut costs). I have expertise in evaluating minute technical details of data, innovating and reinventing systems to embolden C level executive leadership.

I am a Spatial Econometrics and Information systems (Network Scientist) PhD from Purdue Polytechnic Institute of Purdue University. My dissertation projects focused on Collective Intelligence where I applied Data Science and Quantitative modeling (i.e., machine learning algorithms, mathematical and statistical model and network science to examine the formation mechanics/processes of open digital/online platforms and their impact on participant's output (productivity) and choice (contagion).

Dissertation Title; "*Collaboration in Scientific Digital Ecosystems: A Socio-Technical Network Analysis*".

I have,

- Extensive experience in all aspects of Data Science which involve; Data and Graph mining, Machine Learning, Artificial and Business Intelligence, Data and Business Analytics and Statistical and Mathematical Modelling.
- Deep expertise and knowledge in mathematical, statistical, econometrics modeling and development of machine learning and deep learning algorithms and prototypes ready for production.
- Statistical, Machine and Deep Learning techniques: Clustering (SOM-Self Organizing Maps, BM-Boltzmann Machine, AE-AutoEncoders), Classification (ANN-Artificial Neural Networks, CNN-Convolutional Neural Networks, RNN-Recurrent Neural Networks), Natural Language Processing (NLP) and Regression Techniques (Linear, Logistic, Hierarchical, Spatial, Time Series)

Experiences in a variety of computing languages (Python, R, JavaScript, NodeJS, HTML, CSS, MySQL/Hive, Mongo/DB, JQuery), platforms (Windows, Linux) cloud and distributed computing environments (Hadoop/MapReduce, MS Azure & AWS) and Machine learning/Deep learning (SciKit-Learn, NLTK, Keras, Theano, TensorFlow, PyTorch).

EDUCATION

CQF Institute — New York, NY
Quantitative Finance — -Ongoing

Purdue University, Purdue Polytechnic Institute — West Lafayette, IN
PhD in Computer Information Systems and Spatial Econometrics — December 2015

Purdue University, Applied Economics — West Lafayette, IN
MS in Applied Economics — December 2012

Columbia University, Graduate School of Arts and Sciences New York, NY
M.A in Sustainable Development October 2010

University of Nairobi, School of Physical Sciences Nairobi, Kenya
BSc. in Mathematics, Honors November 2001

SKILLS

Programming Languages: Python, R, GAMS, Matlab, Pascal and C++
Software and Other Applications: Stata, CGE, IMPLAN, MySQL, MongoDB, MS Azure, ML Studio, MS Power BI, DOMO, Big Data (Hadoop Architecture/Ecosystem-HFDS, HBase, MapReduce, Hive/Impala/Hue, Scoop, Spark), Linux, Machine Learning, Deep Learning and ArcGIS 9.3
Spoken Languages: Swahili (Fluent), German (Conversational and Reading)

EXPERIENCE

Data Science, Research and Experimental Design Experience

Data Science & AI Consultant New York, NY
Consultant June 2019-

- Consults on Research/Analytics, Data Science, Business and Artificial Intelligence, Machine and Deep Learning and Quantitative Finance projects for Private and Public firms mostly in Developing World.
- Development of news and financial market performance platform/application

Ongoing and Completed Projects.

- *Establishment of Data Science & AI Infrastructure*
- *Teams formation in USA, India & Kenya*

CPG Financial Corp. Services New York, NY
Business Intelligence Manager & Sr. Data Science April 2017-May 2019

- Leads the Business Intelligence Unit team of Research and Data Section at CPG Financial Services Cooperation. My main role is to use data (while applying all Data Science practice tools) to provide business intelligence to Executive Leadership and Business Unit Managers.
- Responsible for developing AI/Data Science/Machine Learning Infrastructure for the Corporation.
- Sits in Data Governance meetings that are responsible for data variety organization, flow and efficient use (in the new Enterprise Data Warehouse)
- Scrum Team Owner of transactional enterprise software development projects in an agile software development framework

Ongoing and Completed Projects.

- *Clients Financial Vulnerability Risk Modeling*
- *Voice of client Sentiment Analysis and Topic Extraction (NLP)*
- *Data Science in-house business units modeling with structured and unstructured (Text heavy data)*
- *Transactional Software development Owner in an Agile Software Development framework*
- *Client/Consumer Demand Projection Models*

Aricent Software Engineering Company Redmond, WA
Senior Technical Leader-Data Science Consultant to Client Microsoft Sept 2016-Dec 2016

- Led a team of Data Scientists, Managers and Developers (Software Engineers) to solve client Microsoft(MS) business problems with data through development of production ready POC (Proof of concepts) and application of all Data Science tools including Machine Learning, statistical and mathematical modeling skills.
- Data Acquisition and Assembly/ETL (Extraction, Transformation and Loading) from various sources including Databases, Warehouses, Web and Flatfiles. *Tools: SQL, MS-Server, R-Server and Python*
- Data Modeling-Machine Learning/Statistical/mathematical modeling and Analysis (Learning

Analytics). *Tools: Python, R, R-Server, MS-Server, MS Azure, MS ML studio.*

- Inference and forecasting (Predictive Analytics) and reporting the actionable insights from the data reporting actionable Insights through interactive business tools (i.e Power BI), Slides and Reports.

Completed POC/Prototypes.

- *Classification of Microsoft X-BOX-Live Consumers based on Business Potential*
- *Apartment Cluster Filters using MS Bing Search API*

Spatial Econometrics, Machine Learning, Information Technology and Economic Policy Consultants West Lafayette, IN
Data Science Consultant to World Bank Jan 2016-Aug 2016

- Consult on broad areas including data science, spatial econometrics, machine learning, information technology and systems and economic policy. *Tools: R, SQL and Stata software.*
- Solve clients' problems from inception to completion applying,
- Data acquisition (extraction, assembly and cleaning) from Data Bases.
- Data preparation (Cleaning, Statistical explorations and transformation-classification/scaling/dimensionality reduction- Transforming Analytics)
- Data modelling and Analysis (Learning Analytics)
- Inference and forecasting (Predictive Analytics)
- Reporting the actionable insights from the data.

Completed Projects: *World Bank DECRG Short Term Consultancies.*

Research Center for Open Digital Innovation West Lafayette, IN
Lead Analyst-Data Scientist May 2014-Dec 2015

- Contributed to the understanding of evolution and optimal operation of online communities and social networks through application of data and graph mining with igraph, spded libraries of R software on structured and unstructured data from nanoHUB.org cyberinfrastructure platform.
- Explored data visually and statistically, classified/clustered and analyzed through machine learning algorithms (mathematical and statistical modelling) using R software libraries.
- Developed both supervised and unsupervised learning machine learning and statistical algorithms to establish patterns of formation, operation and sustenance of software developers' community network using R Software.
- Developed efficiency evaluation application for developer communities through application of stochastic dominance, bootstrapping and exponential random graph statistical models on dynamic (time-series) data.
- Formulated statistical algorithms model to evaluate the impact of developer community network on growth of software developers with supervised Bayesian spatial logistic, fixed effect and interaction statistical and economic models with spdep library of R software.
- Established a social structure and communication channel embedded diffusion model to forecast software usage (revenues) while applying supervised and unsupervised learning models including, bass predictive model, Bayesian spatial probit model and network science.

Completed Projects:.

- *Efficient Management of Information Flow Systems in a Two-Way Market Online Platform*
- *Establishment and Management of Growing Online Communities and Ecosystems*
- *Prediction of Diffusion of Innovation/Products tool based on Advanced Collaborative filters in an Online Setting*

Aquaculture Marketing & IPIA, Purdue University West Lafayette, IN
Lead Analyst-Research Scientist Jan 2011-May 2014

- Computed the economic impact of development projects after devising research questions based on theory, cleaning and classifying data and conducting supervised learning input-output analysis

using IMPLAN and CGE softwares.

- Identified unstructured individual and aggregate firms optimal level of operation based on unstructured data of production.
- Calculated technical efficiency of individual firms using parametric and nonparametric, Stochastic Frontier Analysis and Data Envelopment technical efficiency evaluation models.
- Prepared technical reports, presentations and scientific papers.

Completed Projects:.

- *Simulation of Input-Output systems from (intra and exte) economic shocks using IMPLAN and GCE*
- *Production Optimization Algorithms for Production Units*

Earth Institute, Columbia University New York, NY
Analyst-Research Assistant *in the UN Millennium Village Projects (MVP)* Sept 2009-Nov 2010

- Quantified biannual impacts of 8 Millennium Villages in Africa with statistical modelling after cleaning and analyzing transport and communication data.
- Wrote weekly policy briefs based on the findings, made presentations and attended weekly research advisory meetings.

Kenya Institute for Public Policy Research and Analysis Nairobi, Kenya
Consultant May 2009-June 2009

- Built national energy demand status data through face-to-face surveys using a structured questionnaire.

EATTA/UNEP/GEF Nairobi, Kenya
Consultant May 2008-May 2009

- Estimated the energy demand status and environmental benefits associated with small hydro-power after designing experimental research instrument, collecting, entering, cleaning and statistically analyzing data.
- Prepared technical policy report that advised the project managers on the most cost effective method of selling power and surplus and carbon credits based on the analysis with Stata.

Kenya Development Research Institute-KDRI Embu, Kenya
Research Scientist *in Social-Economics Department* May 2003–Sep 2007

- Designed experimental research and supervised data collection and entry activities.
- Identified periodical impacts of research project based on statistical data analysis.
- Published monthly policy briefs, presentations and periodic technical reports that were used for the center's strategic planning.