



Data Science Africa, Nyeri 2018

Shelmith N. Kairuki [in](#)

& Dina Machuve [in](#)



OUTLINE

- Introduction
- Organizers
- Presenters
- Summer School and Workshop highlights
 - Sensor Deployment
 - TAHMO Weather Stations
 - Bayesian Methods
 - Deep Learning
 - Natural Language Processing (NLP)
 - Model Deployment
- Spotlight talks
- Panel discussions
- Take aways



INTRODUCTION

- Since 2015, Data Science Africa (DSA) has held an annual summer school and workshop.
- The goal of the event has been to train participants on machine learning and data science methods and provide an avenue for researchers to present work demonstrating the application of these techniques to problems relevant in the African context.
- DSA 2018 Nyeri (31st May to 08th June, 2018) was a "***Training of Trainers***" event under the theme "***to-end data science.***"



ORGANIZERS



Ciira wa Maina, PhD
Senior Lecturer
Department of Electrical
and Electronic Engineering
Dedan Kimathi University of
Technology



Dr. Ernest Mwebaze
School of Computing &
IT,
Makerere University





**Neil
Lawrence**
Director of
Machine
Learning
Amazon



Damon Civin
Principal Data
Scientist
ARM



**Thomas G.
Dietterich**
Oregon State
University



Timnit Gebru
Microsoft
Research



**Martin
Mubangizi**
UN Pulse Lab,
Kampala



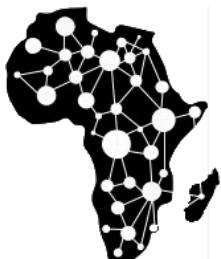
**Jan
Jongboom**
ARM



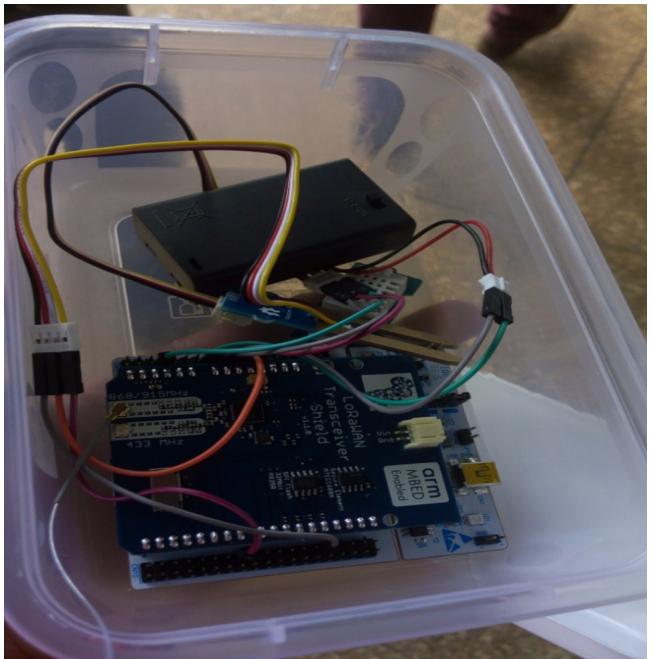
John Quinn
Makerere
University



Sara Hooker
Google



SENSOR DEPLOYMENT



Soil, Humidity and
Temperature
Sensor



Air Pollution
Sensor



Motion Sensor



Trans-African Hydro-Meteorological Observatory (TAHMO)



Weather Stations

- Motivation:
 - Hydro-meteorological data is virtually non-existent.
 - No basis for the determination of how the African climate is changing.
- Designed to provide rainfall, temperature, and other critical data.
- High-quality data freely available to governments, scientists and farmers in real time.
- The link to the lecture can be found [here](#)



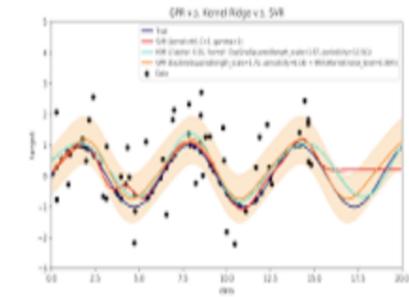
BAYESIAN METHODS



In probability theory and statistics, a **Gaussian process** is a stochastic **process** (a collection of random variables indexed by time or space), such that every finite collection of those random variables has a multivariate normal distribution, i.e. every finite linear combination of them is normally distributed.

[Gaussian process - Wikipedia](#)

https://en.wikipedia.org/wiki/Gaussian_process



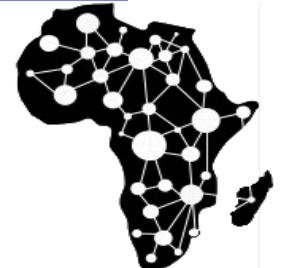
- [Neil Lawrence](#) introduced us to Gaussian Processes and Bayesian Inference.
- Slides to his talk can be found [here](#).



INTRODUCTION TO DATA SCIENCE



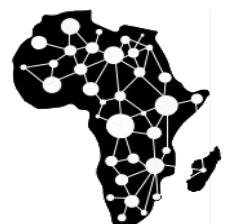
- [Dina Machuve](#) (NM-AIST, Pythontz) introduced the fundamentals of Data Science.
- The key word in “Data Science” is not Data, it is Science.
- What question are you trying to answer with data?
- Data Science = Machine Learning + Statistics
- Slides to her talk can be found [here](#).



COMPUTER VISION AND IMAGE ANALYSIS



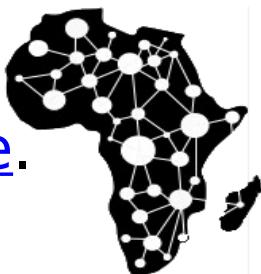
- Ernest Mwebaze (Makerere, UN Pulse Lab) gave a tutorial on the fundamentals of Computer Vision and Image Analysis.
- Cropping is the same as slicing a numpy array
- Image processing inputs an image and outputs an image.
- Computer vision is targeted at generating some understanding or information from an image.
- Slides to his tutorial can be found [here](#).



AIR POLLUTION MONITORING



- Mike Smith (Sheffield University) presented his work on Air Pollution Monitoring in Kampala.
- Data acquired from sensors some deployed in bodaboda
- Air pollution monitoring is real time on mobile phone.
- He gave a Tutorial on Classification Tutorial & Research Software
- Slides for his talk can be found [here](#).



Convolutional Neural Networks



- **Sarah Hooker** is an AI Research Resident at Google Brain in San Francisco.
- Her talk on convolutional neural networks was titled “How do models learn: understanding feature importance in image classification models”
- She is also a volunteer at Delta Analytics with Machine Learning for Good Initiative on Education.
- Slides for her talk can be found [here](#)



DATA VISUALIZATION SCIENCE



- Martin Mubangizi (UN Pulse Lab) gave a tutorial on the fundamentals of Data Visualization.
- Deployed geopandas, pandas, numpy libraries on Ambulance tracking Project in Uganda.
- Using the data here draw a lollipop graph
- Slides for his talk can be found [here](#).



INTRODUCTION TO SPATIAL DATA ANALYSIS



- John Quinn Martin (Makerere, UN Pulse Lab) gave a talk on Spatial Data Analysis for prediction of deforestation.
- Predict based only on the deforestation locations in the previous year
- Used spatial location and some covariates: road density, protected area status, forest cover
- Slides to his talk can be found [here](#)



DEEP LEARNING

Prof. Max Welling from University of Amsterdam & Qualcom gave a talk on Deep Learning

- Deep learning models are vaguely inspired by information processing and communication patterns in biological nervous systems.
- The Deep learning structures discussed were convolutional neural networks and recurrent networks.
- Convolutional neural networks are mostly used for image classification.
- The slides to the lecture can be found [here](#).



MODEL DEPLOYMENT



- **Damon Cavin** (ARM) took us through model deployment.
- He noted that there has been many great tools and tutorials about building machine learning models, but one less talked about topic is how to deploy these models into production in a scalable manner.
- He introduced [clipper](#), which is an open source, model deployment tool.
- Clipper uses cluster (a collection of Docker containers) to serve machine learning models.
- Slides to the talk can be found [here](#).



Interpretability, Bias in Models



- **Julius Adebayo** is an AI Research Resident at Google Brain in San Francisco.
- Interpretability helps to establish trust on the model.
- A big question on interpretability method “How do you know if it's a good explanation or not”
- He gave a tutorial on Deep Learning using TensorFlow.
- Slides and tutorial notebook for his talk can be found [here](#).



Biotechnology: A Natural Home for African Data Scientists



- **Morris Agaba** (FUNAS) gave a keynote on the centrality of DNA, data and biotechnology for agriculture, health and biodiversity.
- “African biologists don’t know maths and data scientists don’t know biology”.
- Overview of ongoing bioinformatics projects:
 - Nubien Ibex genome
 - A gene expression atlas of the domestic goat and comparative analysis in sheep



Interpretability, Bias in Models



- **Kathleen Siminyu** is Lead Data Scientist at Africa's Talking.
- She talked of her career path from a fresh intern to head of Data Science @Africastalking.
- Her work includes building APIs, hosting big data platforms running SPARK and HDFS for processing
- Co-founder of Nairobi [Women](#) in Machine Learning & Data Science



Data integration tool for smallholder farming systems



- **Dennis Mujibi** (NM-AIST, USOMI Limited) gave a keynote titled “Towards 2050: Big data driven decision support in integrated smallholder farming systems”.
- “The world can not produce enough food for itself. Data is the common denominator.”
- USOMI LULU, a real time bench marking decision support tool for small holder farmers
 - Poultry
 - Livestock

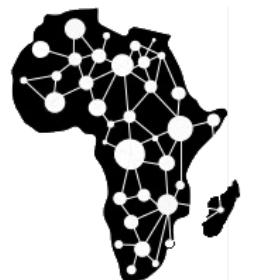


Large Scale Visual Image Analysis



- **Timnit Gebru**, (Microsoft Research, Black in AI) gave a keynote titled,
- “Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US”

- Transfer learning, domain adaptation
- Used Amazon Mechanical Turk (AMT) for crowd sourcing cars
- Google street images, annotation, deep learning



Introduction to Anomaly Detection, density estimation



Prof Tom Dietterich, Oregon State University gave several talks and tutorials on

- Introduction to anomaly detection and density estimation.
- Reinforcement Learning for ecosystem management
- Slides and tutorial notebook for his talk can be found [here](#).



Data Engineering



- **Gen-Tao** is a Data Engineer at ARM gave a key note on Data Engineering and Infrastructure.
- Visualizations tools for exploring data: Kibana, Logstash, Python
- Snapshot to visualize IOT field work data from greenhouse at DeKUT
- A clear data piple is very important for a data science projects.
- Slides and tutorial notebook for his talk can be found [here](#).



Value alignment via Inverse Reinforcement Learning



Join us for this fireside chat with Billy Okal as we discuss **how to handle issues arising when building real-world data-driven AI systems.**

June 9 | 9:00am | Busara Centre

BILLY OKAL

SENIOR RESEARCH ENGINEER AT VOYAGE



- **Billy Okal** is a Senior Research Engineer at Voyage, California.
- He shared on the reinforcement learning involved on self-driving cars
- Key speaker at Meetup at Nairobi Women in ML and Data Science



Data Science in Off-grid Solar

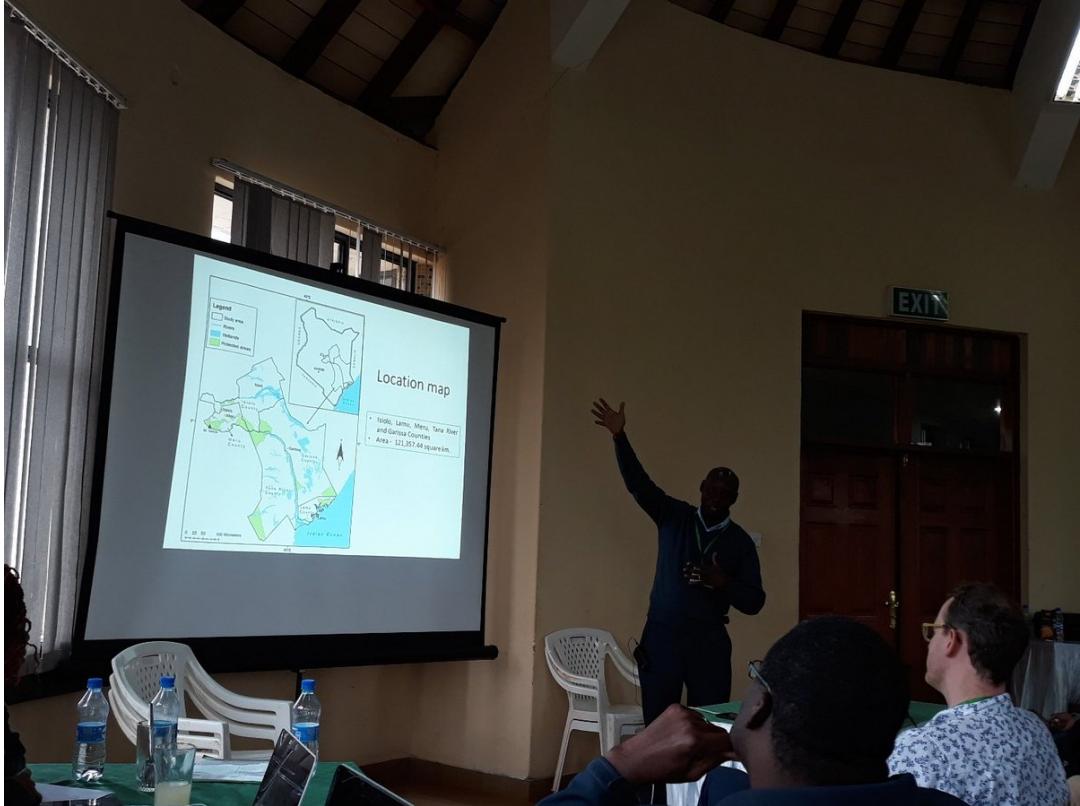


Alex Roussel is a Data Scientist, Fenix International, Kampala.

- Energy utility and financial operations
- Data driven for usage and diagnosis of off solar grid
- Random Forest classifier to predict payment in 30 days following an overdue call
- Diagnostic data sent by GSM chip embedded on panel for identifying sources of poor system performance.



Mapping exposure of cattle to Rift Valley Fever (RVF) Virus



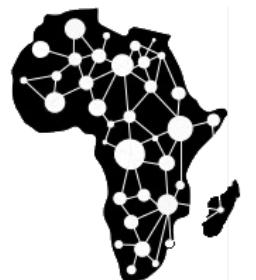
- Prof. Charles Mundia (GIS, remote sensing expert, Dedan Kimathi University) .
- Trajectory cattle movement using GPS collars
- Insights for RVF mapping in Kenya



Intel AI Portifolio



- **Roy Allela, (Intel, Kenya)**
- Intel AI Portifolio - platforms, tools, frameworks, libraries and technology
- Student developer program
- Intel AI Academy (learn, develop, teach, share)
- Intel AI Professor program (tailored curriculum)



SPOTLIGHT TALKS

By students

On Machine learning research work

Ben, Makerere University: Detecting cassava diseases through computer vision

Claire, UTAMU: Intro to Pandas; Text Mining

Neema, NM-AIST: ML for Education, to avoid student dropout



SPOTLIGHT TALKS

By students

On Machine learning research work



Mahlatse, CSIR-SA: analysis of energy utilization and sources in Tshwane city

Mark, Makerere University: Natural Language Processing for robust diseases surveillance in the social web particularly on Twitter

Jared, Dedan Kimathi University; IOT for Smart Agriculture

PANEL DISCUSSIONS

- Data for Public Good
- Building local capacity - Africa Data Science Community



FUN, NETWORKING WITH MENTORS AND FRIENDS



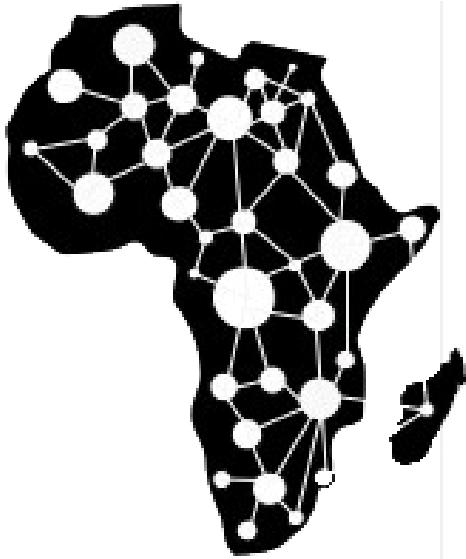
FIELDWORK, NETWORKING WITH MENTORS AND FRIENDS



TAKEAWAYS

- Python is not an option in Data Science and Machine Learning.
- Model building is just not enough, Extensive Research is key.
- Machine Learning projects to solve problems in local context is a breakthrough
- Data Science is all about mathematics, statistics, programming
- Black in AI Workshop in Toronto, Canada – December, 2018
- Meetups are important for career development and motivation





Data for Change in Africa

#DatSciAfrica2018