Conference paper

Data Science Africa – An Initiative to Bridge the Data Science Skills Gap in Africa.

Ciira wa Maina[1*], Mike Smith[2], Ernest Mwebaze[3], Andreas Damianou[2], Martin Mubangizi[4], John A. Quinn[4], Neil D. Lawrence[2]

[1]Dedan Kimathi University of Technology, [2]University of Sheffield, [3]Makerere University, [4]United Nations Global Pulse

*ciira.maina@dkut.ac.ke

Summary

This paper describes an initiative to introduce students in African universities to data science in a workshop setting. We describe the experiences of the inaugural workshop on Data Science in Africa held at the Dedan Kimathi University of Technology in Nyeri, Kenya and the follow-up workshop held in Kampala, Uganda. The workshop consists of lecturers on data science methodology and presentations by researchers on real life application of these methodologies. The workshop was motivated by the realization that students were not adequately prepared to deal with challenges resulting from the data explosion in their areas of interest. The work presented here is an approach to fill this data science skills gap that combines classroom learning, hands on laboratories and presentations by researchers in an intensive one week program.

Introduction

The world is currently experiencing a data explosion with large volumes of data being generated by a variety of devices for example mobile phones and sensors. These data, which have become easier to collect have revolutionized several fields. For example, the ability to sequence genomes has made medical science a data intensive field. Bioinformaticians are now at the forefront of finding cures for the world's ailments (Sander 2000, van El et al. 2013).

In Africa, the mobile phone has made modern communication affordable and accessible to a large segment of the population. In addition to being a means of communication, mobile phones are now used for mobile money which has led to financial inclusion of a large segment of the previously "unbanked" population (Hughes and Lonie 2007). With these developments, a treasure trove of data is available to mobile telephone companies in the form of call detail records and mobility records. These data can be used for social good for example by using mobile phone mobility records to track and monitor the spread of epidemics (Wesolowski et al. 2012).

With all these data becoming available, a new challenge emerges: How do we obtain actionable intelligence from these data? The answer lies in the new and exciting field of data science which is a multidisciplinary field that aims to make sense of the growing amounts of data. At its heart, data science is a quantitative discipline which exploits advances in machine learning and statistics.

In Africa, challenges in areas such as health care, agriculture, energy and wildlife conservation could be tackled much more efficiently if domain experts had the skills and tools to make sense of the data becoming available in their domains. There is a need to train students at various levels to "think with data" and act based on inferences drawn from sound data analysis (Hardin et al. 2015). In this paper we describe an initiative to close the skills gap identified among students in African universities where data science courses have yet to be integrated into the curriculum. We aim to expose the students to both the quantitative and practical skills necessary to be a successful data scientist.

Data Science Africa

The inaugural workshop on data science in Africa (Data Science Africa https://sites.google.com/site/dekutdatascienceworkshop2015/home) was organized at the Dedan Kimathi University of Technology in Nyeri, Kenya from June 15th to 19th, 2015 with the aim of bringing together leading researchers and practitioners working on data science methods or applications relevant to Africa, and to provide training on state of the art data science methods to students interested in developing practical skills. As such it was organized in two parts. First, a three day series of lectures was delivered to approximately 60 students drawn from various departments within the university including computer science, information technology, electrical engineering and geographical information systems. Second, a two day workshop featuring work on real life applications of data science by researchers from around the world.

The second DSA workshop was held in Kampala, Uganda from June 27th to July 1st 2016 (http://www.datascienceafrica.org/). As in the first workshop, the first three days were devoted to training students on various aspects of data science at the Makerere University. The last two days involved presentations and panel discussions by researchers and industry players on applications of data science to various problems. This part was held at the office of the UN Global Pulse Lab in Kampala. As such, there was an emphasis on applications of data science towards the achievement of the sustainable development goals (http://www.un.org/sustainabledevelopment/sustainable-development-goals/).

DSA- The Lectures

The aim of the lectures was to expose the students to the quantitative and software skills needed to apply machine learning methods to data science problems. The lectures were followed by hands on laboratory sessions where students got a chance to experiment with toy and real world data sets. The programming language chosen was Python, and laboratory manuals were presented in the form of Jupyter notebooks. This choice was informed by the wide adoption of Python in the data science community. The lectures topics included:

1. Introduction to Machine Learning and Data Science

2. Introduction to Classification

3. Bayesian Methods

4. Clustering

5. Dimensionality reduction and latent variable models

For DSA 2016 in Kampala, we introduced sessions devoted to interaction with real world data sets and visualizations. For example, participants were introduced to pandas (http://pandas.pydata.org/) and Geopandas (http://geopandas.org/), two python packages for manipulation of data and visualization of spatial data.

DSA- The Workshop

The aim of the workshop was to bring together researchers working in diverse data intensive fields to discuss applications of data science to their domain areas and to expose students to cutting edge research using techniques they had been exposed to during the lectures. The researchers included statisticians, computer scientists, engineers, and bioinformaticians and the work discussed included 1) Use of image classification for malaria diagnosis and plant monitoring (Quinn et al. 2014, Quinn et al. 2011). 2) Analysis of social media and telecommunications data to monitor the spread of epidemics 3) Systems to collect and analyze agricultural data for precision and climate smart agriculture (Williams et al. 2015). 4) Advances in data collection for disaster response (Jain et al. 2015). 5) Data intensive approaches to conservation (wa Maina 2015). A summary of the work presented at DSA 2015 was the subject of an article by Prof. Neil Lawrence in the Guardian newspaper (https://www.theguardian.com/media-network/2015/aug/25/africa-benefit-data-science-information).

Participant Feedback and Assessment

During the first workshop in Nyeri, feedback was informally obtained from the participants via twitter. Using the hashtag #DatSciAfrica, participants shared their thoughts and interacted with one another. A few of the comments included:

1. "it has been an educative week ... Thanks to the organizers"

2. "[This] is one of the most dynamic workshop[s] I've attended."

3. "Great workshop. interacting with Prof. Neil Lawrence yesterday was an honour and a great learning experience"

4. "amazing presentations... Workshop & conference so insightful as I prepare my research in ML"

For the Kampala workshop, a survey was undertaken to obtain feedback from the students. We sought to determine whether the concepts taught during the lectures were well understood and to get the general opinion of the students. The feedback was positive with about 90% of respondents agreeing that the workshop increased their understanding of data science. They also found the integration of lectures and hands-on labs useful. The survey also gave respondents an opportunity to give comments and suggestions. Some of the comments included:

1. "The course structure was well organized and made learning easy to follow and understand."

2. "Educative and insightful!"

3. "... I suggest the summer school period to be extended [to] at least a period of 2 weeks"

4. " … emphasize the need for students to do prior preparatory reading"

The way forward

The third DSA workshop is scheduled for July 17th to 21st, 2017 in Arusha, Tanzania and will follow the same format as the first two workshops. DSA is now an annual event that will provide a platform for researchers to discuss solutions to Africa's problems using data and a chance to educate students on state of the art data science methodology.

Acknowledgements

Competing Interests

The authors declare that they have no competing interests.

References

Hardin, J., et al. (2015). Data Science in Statistics Curricula: Preparing Students to "Think with Data". The American Statistician, 69(4), 343-353.

Hughes, N., & Lonie, S. (2007). M-PESA: mobile money for the "unbanked" turning cellphones into 24-hour tellers in Kenya. Innovations, 2(1-2), 63-81.

Jain, A., Adebayo, J., de Leon, E., Li, W., Kagal, L., Meier, P., & Castillo, C. (2015). Mobile Application Development for Crisis Data. Procedia Engineering, 107, 255-262.

Quinn, J. A., Andama, A., Munabi, I., & Kiwanuka, F. N. (2014). Automated blood smear analysis for mobile malaria diagnosis. Mobile Point-of-Care Monitors and Diagnostic Device Design, 31, 115.

Quinn, J. A., Leyton-Brown, K., & Mwebaze, E. (2011). Modeling and Monitoring Crop Disease in Developing Countries. In AAAI.

Sander, C. (2000). Genomic medicine and the future of health care. Science, 287(5460), 1977-1978.

van El, C. G., et al. (2013). Whole-genome sequencing in health care. European Journal of Human Genetics, 21, S1-S5.

wa Maina, C. (2015). Bioacoustic approaches to biodiversity monitoring and conservation in Kenya. In IST-Africa Conference, 2015 (pp. 1-8). IEEE.

Williams, T. O., et al. (2015). Climate Smart Agriculture in the African Context.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., & Buckee, C. O. (2012). Quantifying the impact of human mobility on malaria. Science, 338(6104), 267-270.