

# Report on the Malawi Data Science Bootcamp 2021

Reuben Moyo  
moyo.r@mzuni.ac.mw  
Mzuzu University  
Mzuzu, Malawi

Stanley Ndevbu  
ndevbu.s@mzuni.ac.mw  
Mzuzu University  
Mzuzu, Malawi

Michael Zimba  
mgmzimba@gmail.com  
Mzuzu University  
Mzuzu, Malawi

Martin Dybdal  
dybber@di.ku.dk  
University of Copenhagen  
Copenhagen, Denmark

Maria Maistro  
mm@di.ku.dk  
University of Copenhagen  
Copenhagen, Denmark

Benjamin Balder Bach  
benjamin@fairdanmark.dk  
FAIR  
Copenhagen, Denmark



Figure 1: Group picture at the Malawi Data Science Bootcamp (MWData 2021), Lilongwe, Malawi, 2021

## ABSTRACT

This report presents and briefly discusses the first Data Science Summer school in Malawi, Africa, which was named Malawi Data Science Bootcamp 2021 (MWData 2021). This event took place at Mzuzu University, Lilongwe ODeL Center in Lilongwe, Malawi, on October 25 - 29, 2021.

## 1 INTRODUCTION

In October 2020, the ACM SIGKDD sponsored the first Data Science Summer school in Malawi, Africa, which was named Malawi Data Science Bootcamp 2021 (MWData 2021)<sup>1,2</sup>. This event took place at Mzuzu University, Lilongwe ODeL Center in Lilongwe, Malawi, from October 25, 2021, to October 29, 2021. The main goal of the event was to build a community that would provide a stimulating and collaborative environment for professionals, students, and lecturers in the field of Data Science. Additionally, the event aimed at improving students' knowledge and skills on data science and empower them to solve problems using data science. Fifty-five participants of various categories: professionals and students at

Masters and PhD levels attended the week-long bootcamp and a total of 17 lecturers and organizers were involved in the event.

The call for participation was issued in two phases; the first phase was for application for attendance and the second phase was a call for scholarship applications. Phase 1 had exactly 100 registrations and we were able to finally select 55 participants among which 50 were awarded scholarships. With the financial contribution of ACM SIGKDD, we supported some of the applicants with travel and accommodation grants. The grants were awarded to students and participants who indicated that they could not afford either accommodation or travel. The event was dominated by participants from Malawi, but also Namibia and Nepal were represented.

## 2 PROGRAM

Malawi Data Science Bootcamp was a single-track event, where all lectures and lab sessions were conducted in a dedicated Computer Lab. Each day was split between lectures and labs. On each day, lecture sessions were held from 08:00 to 12:00, then lunch was served. At 13:00, lab sessions began and went through until 16:45. A plenary session followed and the day was closed. The lectures were delivered in a hybrid format: face to face and online via Zoom. Online lectures were given by presenters who were not able to travel

<sup>1</sup><https://mwdata.science/>

<sup>2</sup><https://twitter.com/mwdatabootcamp>

to Malawi due to Covid-19 travel restrictions. However, this did not affect the programme, as all sessions were delivered seamlessly.

The MWDData bootcamp was officially opened by Associate Professor Chomola Mikeka, Director of Science, Technology and Innovations, in the Ministry of Education in Malawi, followed by Mr Yonamu Ngwira, the University Registrar and Dr Artz Luwanda, the Dean of the Faculty of Science, Technology and Innovation at Mzuzu University. On the closing day of the bootcamp, participants were awarded a certificate of participation. Then, on the evening of the final day, participants and local organizers gathered for a dinner to socialize and build connections.

Next, we summarize the programme of each day. Day 1: The event kicked off with an official opening ceremony and then a keynote lecture on the importance of datasets for Machine Learning (ML) and data science, delivered by Dr. Amelia Taylor of the Malawi University of Business and Applied Sciences, Malawi. Wiza Msuku of Elizabeth Glaser Pediatric Foundation delivered a lecture and a lab session on data pipeline and understanding the basics of the Extract, Transform, and Load (ETL) pipeline. A lecture and lab session on Web Scraping was delivered by Ralph Tambala of Malawi University of Science and Technology, Malawi. Day 2: The day started with a lecture on machine learning algorithms with their applications by Akuzike Banda followed by a lecture and lab session on Linear Regression, facilitated by Winnie Mkandawire. Day 3: On the third day, Tiwonge Banda introduced participants to the Support Vector Machine (SVM) algorithm through a lecture and lab session. Later in the day, Richard Munthali presented a lecture on mathematics and statistics concepts on ML and data science, with a focus on the healthcare domain. Day 4: The fourth day continued with Maria Maistro delivering a lecture and lab session on databases for data science. This was followed by a keynote lecture by Rachel Sibande on use cases for big data for development use cases. Later in the afternoon, the participants were divided into groups and a short project was assigned to each group. Day 5: The week ended with a lecture on big data by Michael Zimba. Then participants presented solutions and findings on the group projects assigned on day 4. The day was then closed by the awarding ceremony, where participants were awarded certificates of participation. In the evening, participants and local organizers gathered for a dinner to socialize and build connections.

### 3 EVALUATION AND OUTCOMES

Before and after the event, participants received online evaluation questionnaires. Next, we summarize some of the main insights and lessons learned from the event.

#### 3.1 Pre-event Evaluation

A pre-event evaluation form was shared with the participants prior to the event in order to get information on the expectations and background of the participants. Feedback from 34 participants out of 45 was received, representing a 75.6 percent response rate. The feedback from the survey assisted the organizers and facilitators to adjust the event based on the expectations of the participants. The feedback also assisted the facilitators in developing lectures and labs of appropriate levels. Several questions were asked on expectations and background, among them, of significant interest was the

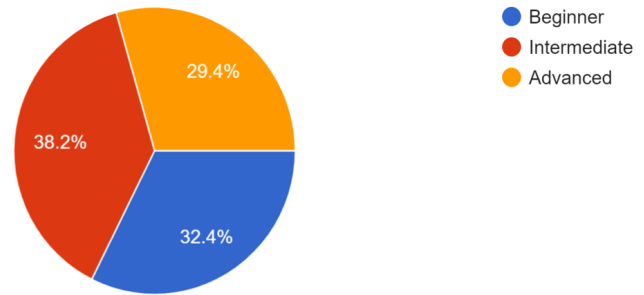


Figure 2: What level of training are you looking for or do you expect?

feedback on what level of training the participants expected (see Figure 2). The results in Figure 2 show an almost equally balanced distribution: 29.4 percent expected an advanced level of training, 38.2 percent required an intermediate level, and 32.4 percent expected a beginner level of training. Based on this, lecturers were encouraged to prepare exercises with different levels of difficulty, so that participants could choose exercises based on their level of background knowledge.

#### 3.2 Post-event Evaluation

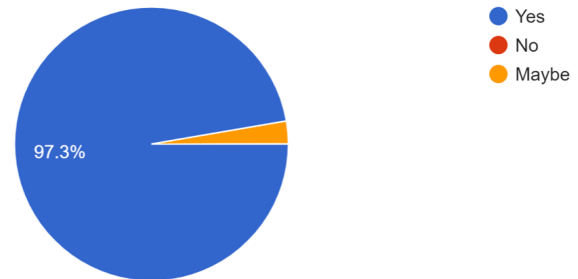
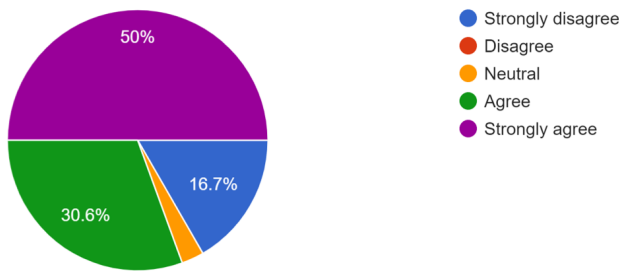


Figure 3: Did the bootcamp meet your expectations?

A post-evaluation survey was conducted at the end of the bootcamp and the responses were extremely positive. Overall 100% of respondents indicated that the content of lectures and keynotes had been relevant to their professional work. In terms of students' expectations, 97% indicated that the bootcamp met their expectations, with only 3% answering "maybe" (see Figure 3).

We observed that the hybrid format with some remote facilitators had a negative impact on students' engagement and interaction, with around 17% of participants lacking participation to some degree (see Figure 4).

Overall, all the participants indicated the need for the event to be recurring. Participants suggested some modifications: (1) increasing the number of participants; (2) working on group projects that could



**Figure 4: Was the training interactive and engaging?**

lead to publications; and (3) increasing the length of the event, e.g., a 2-weeks bootcamp.

### 3.3 Takeaways and Outcomes

Participants were given access to a common chat platform, a self-hosted Matrix instance. Because the platform was new to most, people had to spend some time becoming familiar and in some cases also fetching an Android client. This self-hosted platform promoted safe sharing of photos and exercises, but in the future we should seek to capitalize more on having our own instance. The Matrix instance is still up and running, such that participants can continue to network and for future events to continue.

Secondly, the Covid-19 pandemic put a considerable strain on organizing the event. Dates were postponed several times and we had many discussions related to travel restrictions which were preventing abroad participation, both for facilitators and participants. Covid-19 regretfully became the direct reason for 4 facilitators' cancellations, 3 of which would have traveled from abroad. Remote presentations affected the interactivity and engagement of participants and facilitators. Despite that the engagement took a hit from the fatigue of perhaps too many remote lectures (3 in total), we also gained valuable experience from the setup and feel optimistic that it can be reused in the future.

Thirdly, a digital infrastructure was also set up in order to manage registration and mass-communication with participants. We

experienced a lot of problems with mass emailing until we decided to set up our own (free) SMTP instance and a self-made form for emailing participants with information leading up to the event. This whole setup remains intact for a future event. As organizers, we strongly encourage that a registration system is built up to handle the diverse needs of participants and speakers or facilitators: dietary restrictions, accommodation, allowances, travel distances, registration status (accepted/rejected/waiting list) and even t-shirt sizes. All such details should be gathered systematically, as well as sending out reminders. If not, the communication can easily become overwhelming. We managed to get through the preparation of the event with relatively little manual communication and most practical aspects handled through online registration forms, in which each participant had a unique token for entering and editing their registration details.

## 4 CONCLUSION AND REMARKS

The Malawi Data Science Bootcamp 2021 (MWData2021) provided a platform for upcoming and experienced data scientists around the region to build a community of practice in data science. The experiences and the lessons that were shared by both facilitators and participants provided the much needed build up to the significance of data science in decision making. The response from both participants and facilitators to the bootcamp was highly overwhelming and the interest to continue with this event is very high. We hope that the recurrent event will continue to inspire and build long-lasting bonds and projects for the ones involved.

## ACKNOWLEDGMENTS

We acknowledge the financial support that was provided by ACM SIGKDD through the "Community Impact Program" call. The support boosted participation through travel grants, accommodation, and the organization of the event. We also acknowledge the technical support provided by the Computer Science Department of the University of Copenhagen (DIKU) and FAIR Denmark. In addition, we acknowledge the support of the host institution, Mzuzu University and the Directorate of Science, Technology and Innovation in the Malawi Ministry of Education for their contributions to the opening ceremony.