tralac e-Learning Course 2020
Introduction to Data Science

4-8 May 2020
Facilitator: John Stuart, tralac Associate

Introduction and motivation

The deluge of data generated by online activities and activities involving devices and the internet of things (IOT) has led to the growth of a new area of specialisation entirely dedicated to the effective usage of this data. 'Data Science' refers to the practice of acquiring, analysing and communicating insights from data, with the primary purpose of generating actionable insights to the solving of problems. The obvious example is that of an e-commerce business wishing to acquire insights on how better to market products to its customers, but the same tools are used to generate solutions to medical, biological, security-related and climactic problems too (among others). Indeed it is safe to say that at the time of writing, new applications for the techniques of data science are being discovered and applied.

While generally not requiring the technologies related to 'big data', trade analysis can benefit substantially from the other cornerstone of the data science field – statistical learning. These techniques (known as 'machine learning' when applied on modern high powered computers) can reveal accurate predictions and inference around many of the problems that trade analysts are required to solve.

The Trade Law Centre (tralac) is an NPO focussed on trade, trade and industrial regulation and economic integration in Africa. In order to help trade analysts and economists come up to speed with the new technology, tralac is offering a one-week introductory e-Learning course to the techniques and methods of data science. This course will help orientate students around the new practices, tools and terminology. The course will be *applied* in the sense that the students will be required to complete exercises during the learning process, as well as a small project with results presented on the final day. The course will introduce, but not require coding for completion, rather the data science workflow will be completed using the **Knime Desktop Analytics Platform** (www.knime.org).

Requirements

- This course is intended for practitioners who already work with, and analyse data using tools such as Microsoft Excel. No pre-training is provided on MS Excel or on basic data analysis.

- All students are required to complete an online evaluation and submit a motivation. In their motivation, students should describe how they currently work with data, why they wish to learn about Data Science techniques and what they intend to get out of the course.
- All students must possess their own Windows laptop computer, which should be a recent model and should have at least 3G of RAM and around 2G of usable storage space for software and data. Students need to have administrator rights on this laptop (ie the right to install new software).