![tralac | trade law centre logo]

tralac Training Week 2022: Introduction to Data Science e-Learning Course

09 – 13 May 2022
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, agricultural, security-related and geoscientific 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 NPC 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

1

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.com).

Programme

Most sessions are a combination of teaching via online webinar presentations and interactive webinar sessions covering applied exercises using Knime. The primary deliverable by participants is the project, which is based on a tralac case study. Details of the project are given below.

Day 1		
	Morning Session 1	**Presentation 1**: Introduction to data science, Knime desktop platform and overview of course objectives
	Morning Session 2	*Lab 1*: Knime: installing, setting up and basic orientation
	Afternoon Session 1	**Presentation 2**: Extract-Transform-Load (ETL) with Knime
	Afternoon Session 2	*Lab 2*: ETL with Knime exercises
Day 2		
	Morning Session 1	**Presentation 3**: ETL part 2 with Knime
	Morning Session 2	*Lab 3*: ETL part 2 exercises
	Afternoon Session 1	**Presentation 4**: Data visualisation with Knime
	Afternoon Session 2	*Lab 4*: Data visualisation with Knime
Day 3		
	Morning Session 1	**Presentation 5**: Machine Learning: Regression and Artificial Neural Network
	Morning Session 2	*Lab 5*: Machine Learning with Artificial Neural Network
	Afternoon Session 1	**Presentation 6**: Machine Learning: Decision Trees
	Afternoon Session 2	*Lab 6*: Machine Learning with Decision Trees
Day 4		
	Morning Session 1	**Presentation 7**: Machine learning with Naïve Bayes
	Morning Session 2	*Lab 7*: Machine learning with Naïve Bayes
	Afternoon Session 1	**Presentation 8**: Case Study: Predicting Intra-Africa Trade Convergence part 1
	Afternoon Session 2	Case Study: Predicting Intra-Africa Trade Convergence part 2/Own Projects Work
Day 5		
	Morning Session 1	Own Projects Work
	Morning Session 2	Own Projects Work
	Afternoon Session	Presentation of class projects

eLearning approach

Participants attend the twice-daily webinar-based presentations then attempt the labs thereafter, again in the context of a live webinar. The facilitator will be available daily in the interactive webinar sessions, to take and answer questions. The final presentation will discuss the class project based on a tralac paper – *Introducing Data Science Techniques for Trade Analysis Applications in Knime*.

Participants will work on the project on the Thursday afternoon until the Friday lunch time. They will then submit their final drafts by email and will be asked to summarise their findings in the group online meeting on the Friday from 2-3pm.

All presentations, labs and supporting documents will be available to download via the Classe365 learning management system.

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 participants are required to complete an online evaluation and submit a motivation. In their motivation, participants 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. The evaluation form is located here: http://tiny.cc/9jsquz
- All participants must possess their own Windows laptop computer, which should be a recent model and should have at least 4GB of RAM and around 2GB of usable storage space for software and data. Students need to have administrator rights on this laptop (i.e. the right to install new software).
- In order to participate, the free Zoom webinar platform is required to be installed on the participants' laptops; access to Classe365 is browser-based and will be provided by tralac staff.

Project

The completion of this course requires the completion of a hand-in project that is based on the material covered in the course. The background reading is the course work plus this tralac paper:

> Stuart, J. 2019. *Introducing Data Science Techniques for Trade Analysis: With Applications in Knime*. tralac Working Paper No. S19WP07/2019. Stellenbosch: tralac

This paper will be distributed to students, but can also be downloaded from the tralac website (https://www.tralac.org/publications/article/14190-introducing-data-science-techniques-for-trade-analysis-with-applications-in-knime.html)

Project description

Can you develop a predictor of intra-African trade convergence? That is, can you predict whether an African country is exporting relatively more over time – that is defined as intra-African trade convergence.

Tips

Calculate change in intra-Africa trade using 10-year Trade Map data, 2008 to 2018. Make this a RELATIVE change, i.e., have exports from the African country to other African countries risen RELATIVE TO TOTAL EXPORTS over the 10 years? This will be provided to participants to download ('Intra-Africa Converging'.xls)

The World Development Indicators (WDI) has multiple indicators of trade, economic and social variables for most African countries spanning many years. Also calculate the 10-year change for these variables (whichever you select) so that, for example:

- Increase in FDI => predicts intra-Africa convergence?
- Reduction in border friction => predicts intra-Africa convergence?
- Reduction in employment in Agriculture => predicts intra-Africa convergence?
- Increase in financial inclusion => predicts intra-Africa convergence?
- Increase in mobile penetration => predicts intra-Africa convergence?

Interested?

Please contact: Info@tralac.org (by 28 April 2022)