

# DEMOCRATIZING DATA SCIENCE EDUCATION FOR THE NON-SCIENCES



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

Nkosikhona Msweli

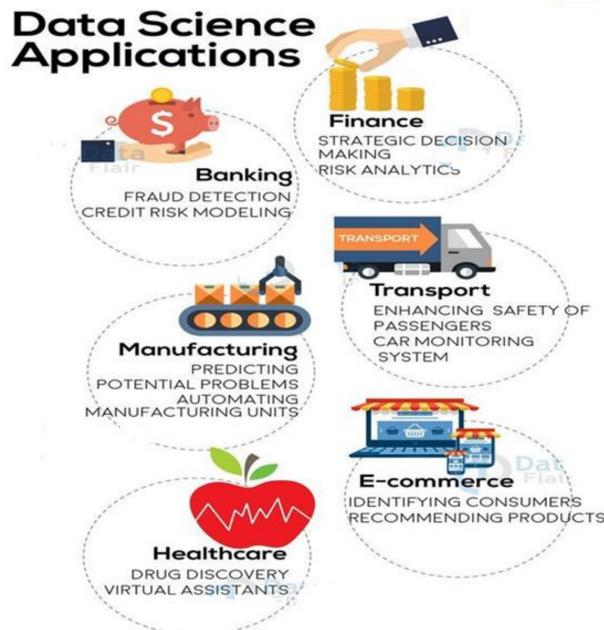
UNIVERSITY OF PRETORIA



## Introduction

Data is today described as a new and desirable and organisations increasingly require data science skills to harness this commodity. The proliferation of data and the affordance of modern technologies has triggered many organizations to contemplate data science. However, data science skills are scarce, and even more challenging is that data science education and training programs are inconsistent and adequate.

Data science or the “science of data” is one of the emerging ICT specialist disciplines that have attracted many scholars in information systems (IS). Data science tasks range from the technical understanding of business problems to coming up with algorithms to generate knowledge (Wang et al., 2021). Below are some examples of how data science is applied in different industries.



## Problem Statement

The interdisciplinary nature of data science raises concerns about competencies/skills, resources, and commitment within the education and training space. The main challenge is that there are not enough data science skills available, especially in non-sciences. This is due to lack of programs focusing on data science education for non-sciences. As it stands, the gap between organisational demand and supply of data science cannot be filled immediately. Data science is fast-paced, and the existing pedagogies cannot adapt quickly to fill the gap. This study plans to investigate how micro-credentials may support data science education for non-sciences to close the skills gap.

## Research Question

How can data science education be democratized for the non-sciences using micro-credentials?

## Purpose of the Study

The study intends to investigate how to democratize data science education for the non-sciences, with micro-credentials as the ideal mechanism.

## Methods

### Design Science Research (DSR)

To solve field problems by designing artefacts that practitioners can apply rather than filling the knowledge gap

### Process Model: Elaborated action design research (e-ADR)

The e-ADR (Mullarkey and Hevner, 2019) serves as an extension and alteration for the original ADR (Sein et al., 2011). This theory is beneficial by providing a well-structured process model that is inclusive of action research and design science research (Sein et al., 2011). The study will follow the four (4) phases as

1. Diagnosis
2. Design
3. Implementation
4. Evolution

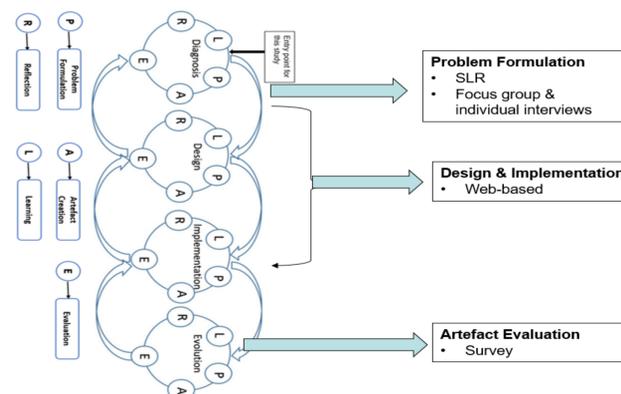
### Critical Realism

The study argues for critical realist informed sociotechnical network perspective, that it offers an opportunity to conceptualize a framework that considers socio-technical issues, humans, organisations, and technology (material) dimensions and can provide the basis of a more practical approach to researching and building data science education artefact.

## Data Collection

Two-phased data collection

- Phase 1 - Focus Group (Problem Formulation)
- Phase 2 – Artefact Evaluation



## Population Sample

- **IS/IT and data science lecturers (public educational institutions)** - course structures and development, how to embed technology, and teaching pedagogies.
- **The university executive management** - perspective on policy such as recruitment of students and resource allocation.
- **Employers (ICT sector)** - how data science roles are filled within their organisations, the role they can play in data science education?
- **Data Scientists** - Insights on qualifications required, challenges they face in the field

## Expected contribution

### Theoretical Contribution

- The study will offer guiding principles towards implementing data science education micro-credentials for non-sciences

### Methodological Contribution

- Lessons from applying the recent e-ADR process model by Mullarkey and Hevner (2019)

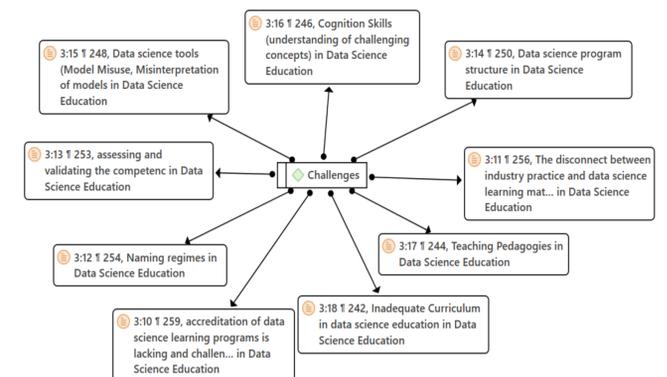
### Practical Contribution

- This study proposes a framework towards democratizing data science education for non-sciences
- Case study on non-science domain micro-credential
- The study will further propose a digital prototype for hosting data science education micro-credentials

## Work in Progress

Data Science Education - A Systematic Literature Review

The study presents a review on data science education research. The main aim is to identify the opportunities for and challenges in data science education



## Bibliography

1. Demchenko, Y. et al. (2019) ‘EDISON Data Science Framework (EDSF) Extension to Address Transversal Skills required by Emerging Industry 4.0 Transformation’, in 2019 15th International Conference on eScience (eScience). IEEE, pp. 553–559.
2. Demchenko, Y., Communiello, L. and Reali, G. (2019) ‘Designing customisable data science curriculum using ontology for data science competences and body of knowledge’, in ACM International Conference Proceeding Series. Association for Computing Machinery, pp. 124–128. doi: 10.1145/3322134.3322143.
3. Mullarkey, M. T. and Hevner, A. R. (2019) ‘An elaborated action design research process model’, European Journal of Information Systems ISSN:, 28(1), pp. 6–20. doi: 10.1080/0960085X.2018.1451811.
4. Witmer, J. (2021) ‘Inclusivity in Statistics and Data Science Education’, Journal of Statistics and Data Science Education, 29(1), pp. 2–3. doi: 10.1080/26939169.2021.1906555.
5. Radovitsky, Z. et al. (2018) ‘Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis’, Journal of Supply Chain and Operations Management, 16(1), pp. 82–101.