Minutes of MSc Dissertation Supervision Meeting

Student: Grace Wangui
Supervisor: Dr. John
Date: 24 June 2025
Duration: ~30min
Format: Physical meeting

1. Meeting Purpose

To finalize the analytical direction of Grace's MSc dissertation project and evaluate methodologies that align with CSA Africa's goals and the available dataset.

2. Project Focus Confirmation

- **CSA Africa's research objectives** were restated:

    - **(1)** Evidence the existence of a programming/digital skills gap.

    - **(2)** Identify the **causal factors** behind this gap (e.g., infrastructure, motivation, socio-demographics).

- Grace suggested predicting a **composite "skill gap" score** using indicators such as:

    - Self-rated competence.

    - Time spent coding.

    - Prior training/workshop attendance.

3. Analytical Approaches Discussed

Agreed Strategy: Compare 3 Modelling Approaches

Grace will explore **three distinct methodologies** to model and analyze the skill gap:

A. Traditional Machine Learning

- Use logistic regression (or similar).

- Focus: Predict skill gap score using background features.

- Baseline model for comparison.

B. Bayesian Probabilistic Modelling

- Use **PyMC** for probabilistic inference.

- Construct an interpretable model with uncertainty estimates.

- Could model the composite skill gap target variable as a **latent variable** influenced by observable indicators.

C. Causal Inference

- Use **DoWhy** (or PyMC's causal module).

- Build a **causal graph** to model directional relationships between features and skill gap.

- Enables **counterfactual analysis** (e.g., "What if access to internet changed?").

Supervisor emphasized this route is **technically challenging but rewarding**, and Grace is encouraged to try all three, then scale if needed.

4. Discussion on Clustering

- Clustering (e.g., K-means or Latent Class Analysis) was discussed.

- Consensus: If used, it would serve as a **preprocessing step** to create new categorical variables (e.g., learner archetypes).

- Not a core method for this phase.

5. Data Considerations

- No immediate need for additional data collection.

- Ethical clearance for new data may not be feasible in time.

- Grace's current dataset is **sufficient**, though some features may need engineering (e.g., deriving structured features from text).

6. Tools & Resources

- **Bayesian modelling**: PyMC

- **Causal inference**: DoWhy

- Supervisor will share relevant **tutorials, links, and notebooks**.

7. Next Steps and Action Items

Task	Responsible	Deadline
Finalize composite skill gap variable design	Grace	Before next meeting
Explore software (PyMC, DoWhy) and install	Grace	Before next meeting
Conduct preliminary modelling trials	Grace	Ongoing
Touch base with Sofiat (CSA) for alignment	Grace	Before next meeting
Supervisor to send tutorial resources	Dr. John	ASAP
Prepare discussion/comparison outline of three methods	Grace	By next meeting