Msc Project Ideas based on CSA

Option 1: Mapping a Barriers–Needs Matrix for African Tech Learners Using Probabilistic Models/ Latent Class Analysis

Research Problem / Motivation

CSA aims to empower young Africans with computing skills but needs a *systematic, data-driven way to understand the major challenges learners face* and what kind of support would help them thrive >> targeted help?

Currently, learner challenges are **qualitatively captured / understood** (e.g., lack of laptops, no stable electricity), but there's no structured or predictive mapping of which learners need which support. This limits CSA's ability to provide **targeted**, scalable interventions.

This project will asks:

- Can I statistically uncover distinct learner profiles based on the types of barriers they face?

- Can I model what kinds of support are most relevant for each profile?

Core Objectives

1. Identify latent learner groups based on barriers to participation and success in computing.

2. Characterize each group's needs, goals, and demographic profiles.

3. Build a barriers vs needs matrix that informs CSA on what types of support work best for whom.

4. Recommend targeted interventions for each learner group.

Methodology <refine this>>

1. Data Preprocessing
- Clean and encode categorical survey responses (barriers, demographics, goals).
- ***What to do about the open ended / text answers? Vector embeddings+similarity???

2. Latent Class Analysis (LCA)
- Use LCA to uncover latent learner groups with similar patterns of barriers
- *** How to correctly determine the number of classes to go for???
- ***Tools to do LCA??

3. Cluster Profiling
For each class from step 2, describe the barriers, demographics, and their goals . Maybe motivation too?

4. **Mappings**

- Analyzing what types of support each class **requests** / **needs**.

- Building a **Barriers vs Needs Matrix** that:

  - Rows = barrier profiles

  - Columns = support types

  - Cell = proportion or predicted benefit from support (e.g., 80% of Class A need mentorship)

    OR use K Means for clustering instead of LCA ?

5. Evaluation

- Use logistic regression to understand which features drive belonging to each class (other feature engineering techniques?)
- class separation (entropy, silhouette scores

Outputs- barriers vs needs model, learner profiles, dissertation

Option 2: Profiling Potential Learners Archetypes Using Unsupervised Learning and Network Analysis

Problem Statement / Motivation

CSA Africa needs to understand who their learners are in a data-driven way and not just demographically, but in terms of their **attitudes, challenges, motivations, and resources**. Why? To improve on their strategies and to maximize their impact

This project will **extract meaningful learner archetypes** : I will use

- Unsupervised machine learning

- Network-based clustering

- Possibly graph embeddings or community detection

Why do this?

- Tailor their curriculum to specific learner types

- Target mentorship and hardware support more effectively

- Design **persona/ motivation-driven journeys** to tech careers

Objectives

1. Cluster the potential learners into their **profiles (**archetypes**)**
2. Visualize how these archetypes **relate** to each other e.g., via similarity graphs.

3. Characterize each archetype by its motivations, barriers, support needs, and possibly confidence from the open ended questions.

Methodology

1. **Data Preprocessing**
   - Encode categorical)

   - Normalize

   - Deal with missing values if necessary

   - Select key features:

     - Motivation (open text or multiple-choice)

     - Confidence

     - Access (device, internet)

     - Prior experience

     - Goals (learning structure, career, time)

2. **Unsupervised Clustering Approaches**
   Choose 2/3 and compare

3. **Evaluate 3 above**
4. **Network Analysis** - Build a **similarity graph** between learners:

   - Nodes = learners

   - Edges = similarity above threshold (e.g., Jaccard or cosine similarity)

Use networkx from webscience class for visualization

5. **Archetype Profiling**

For each profile:

   - Identify key characteristics (mean feature values, mode answers)

   - Name and describe archetypes (e.g., "Tech-Curious Beginners", "Reskilling Adults", "Confident Coders with No Support")

Outputs
Learner archetype labels, Cluster evaluation, Similarity graph, dissertation

<— For both I will do literature review for tech pipelines/ opportunities/ barriers/ motivations/ current solutions in Africa—>