Research can be classified by SDG in **concatenating** abstracts with descriptions of SDG **targets** and **balancing** training samples across the 17 SDGs.

An SDG classification tool for South African research

Group UniMindz: Lamont Theron, Khutso Sepuru

1 Problem

- Policy formulation and implementation as well as research funding should be informed through evidence
- Improving the SA SDG Hub's research classification platform can enable this

Our Research Questions:

- Can classification accuracy of research articles be improved relative to the 17 SDGs?
- Can articles be classified to any of the 169 SDG targets?

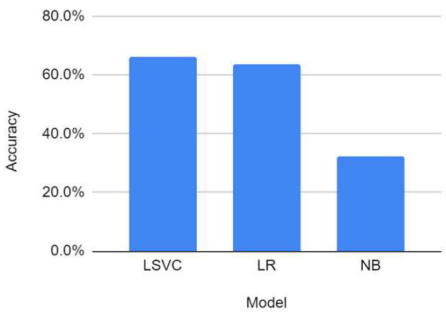
2 Approach

- Scraped articles and SDG tags from SA SDG Hub site
- TFIDF of abstracts and target descriptions concatenated
- Resampled to balance SDGs

Our Models:

- LSVC: Multilabel Linear Support Vector Classifier
- Multiclass Linear Regression
- Multiclass Random Forest

3 Results



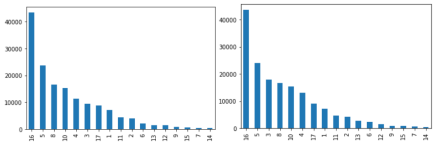
- Accurate if all SDGs per sample match
- Classifying targets had good cross-validation F1=0.71, but bad test F1=0.04

4 Next Steps

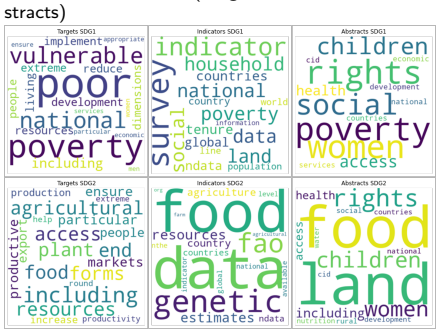
- Include SGD indicator metadata (concatenate or sample)
- Resample to balance training data across SDG targets
- Train multilabel models on One-Hot-Encoded SDG targets
- Deploy multilabel models to streamlit

Extra figures

Frequency by SDG (Training vs Training + Scraped):

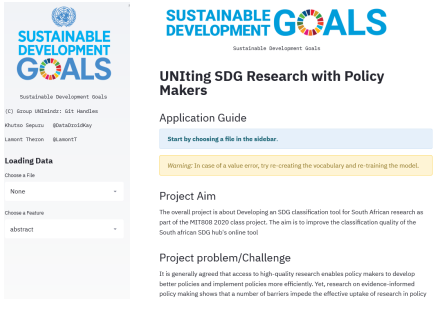


Word clouds SDG1-6 (Target vs Indicators vs Abstracts)



Abstract words not as prominent in Definitions:
SDG1: 'rights', 'woman', 'children'
SDG2: 'land', 'rights', 'woman', 'children', 'health'
SDG3: 'rights', 'woman', 'children','education'
SDG4: 'rights', 'woman', 'committee', 'health'
SDG5: 'rights', 'committee', 'children'

Streamlit interactive web app:



Abbreviations

SDG: Sustainable Development Goals

TFIDF: Term Frequency Inverse Document Frequency

Department of Computer Science
Faculty of Engineering, Built Environment and Information Technology
Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Capstone Project - MIT 808

Course Coordinators:
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)
Abiodun Modupe (abiodun.modupe@cs.up.ac.za)