Classifier may exhibit **bias** due to **class imbalance**. Remedies include **user validation** and **synthetic text generation**.

Automatic Fuzzy Classification of Abstracts as per UN SDG's

Laing Lourens, Davod Khan

1 Intro

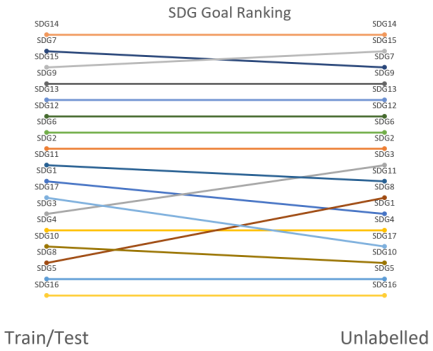
- UN Sustainable Development Goals prioritise investments
- Progress estimated by classification of published research
- Goal: improve accuracy with which research articles are classified
- Goal: improve digestability of these results

2 Methods

- $N_{labelled} = 90216$
- Train-Test split 80:20
- Preprocessing steps:
  - language identification
  - filter stopwords
  - filter punctuation
  - tokenization
  - GloVe embedding vectors
- LSTM Neural Network model

3 Results

- $N_{unlabelled} = 22448$



- Despite strong performance:
  - Rankings closely match training data set
  - This is potential evidence of bias in the model

4 Future Work

- Validate model outputs with `streamlit` tool
- Address class imbalance with synthetically generated articles

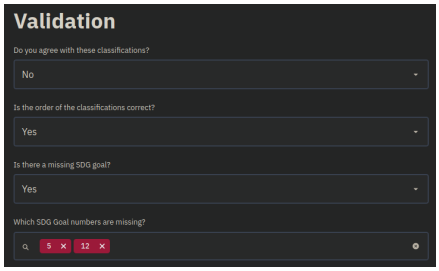
Extra figures

Performance Metrics

- *precision* : 88.13%
- *accuracy* : 97.25%
- *recall* : 82.67%
- *loss* : 7.48%

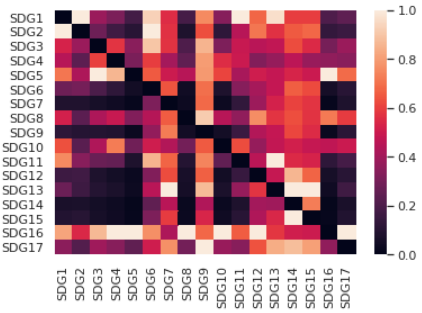
`Streamlit` Application

- Goals, Targets and Indicators Exploration Tool
- Text Classification and Validation Tool



Coincidence Matrix

- This normalised data structure provides a basis to build a recommendation engine for similar goals during user validation



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)

Scan me