A machine-learning approach for a CRIS research outputs’ SDG classifications

euroCRIS Strategic Membership Meeting 2023, Pamplona, Spain

António Luís Lopes IT-IUL (Iscte)
Catarina Roseta-Palma BRU-IUL (Iscte)
Ana Simaens BRU-IUL (Iscte)
AGENDA

01 About Iscte and our CRIS

02 SDGs in the CRIS

03 Applying Machine Learning

04 Evaluation

05 Conclusions
About Iscte and our CRIS
The Iscte - University Institute of Lisbon was founded in 1972 as one of Portugal’s first modern universities, with the central aim of studying labour and social dynamics in a rapidly changing world.

Since its establishment, the university has expanded its disciplinary reach into five schools in the following areas: Business, Sociology and Public Policy, Technology and Architecture, Social Sciences and Humanities, and Applied Digital Technologies.
The work on Iscte’s CRIS started in 2010 as part of the development of an evaluation system for the University. In 2013 it spun off as its own individual system separated from the evaluation system.

From then on, it evolved to gather all kinds of relevant scientific and research information and to focus on internal and external integrations to avoid any duplicate efforts in collecting or providing data to the stakeholders.
02

SDGs in the CRIS
Sustainable Development Goals
Iscte’s Publications and Projects that contribute to the SDGs

1. NO POVERTY
   Nr. publications: 246
   Nr. Projects: 16

2. ZERO HUNGER
   Nr. publications: 60
   Nr. Projects: 4

3. GOOD HEALTH AND WELL-BEING
   Nr. publications: 1126
   Nr. Projects: 125

4. QUALITY EDUCATION
   Nr. publications: 1809
   Nr. Projects: 317

5. GENDER EQUALITY
   Nr. publications: 565
   Nr. Projects: 61

6. CLEAN WATER AND SANITATION
   Nr. publications: 55
   Nr. Projects: 6

7. AFFORDABLE AND CLEAN ENERGY
   Nr. publications: 168
   Nr. Projects: 18

8. DECENT WORK AND ECONOMIC GROWTH
   Nr. publications: 1300
   Nr. Projects: 138

9. INDUSTRY, INNOVATION AND INFRASTRUCTURE
   Nr. publications: 1435
   Nr. Projects: 74

10. REDUCED INEQUALITIES
    Nr. publications: 1779
    Nr. Projects: 238

11. SUSTAINABLE CITIES AND COMMUNITIES
    Nr. publications: 1468
    Nr. Projects: 94

12. RESPONSIBLE CONSUMPTION AND PRODUCTION
    Nr. publications: 691
    Nr. Projects: 16

13. CLIMATE ACTION
    Nr. publications: 436
    Nr. Projects: 28

14. LIFE BELOW WATER
    Nr. publications: 64
    Nr. Projects: 5

15. LIFE ON LAND
    Nr. publications: 195
    Nr. Projects: 13

16. PEACE, JUSTICE AND STRONG INSTITUTIONS
    Nr. publications: 1103
    Nr. Projects: 113

17. PARTNERSHIPS FOR THE GOALS
    Nr. publications: 261
    Nr. Projects: 17
Sustainable Development Goals
Iscte’s Publications and Projects that contribute to the SDGs

Quality education
Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

Apply filter: Iscte - Instituto Universitário de Lisboa

Publications (1859)

<table>
<thead>
<tr>
<th>Type</th>
<th>Title</th>
<th>Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific journal paper</td>
<td>Tackling regional skill shortages: From single employer strategies to local partnerships</td>
<td>04</td>
</tr>
<tr>
<td>Scientific journal paper</td>
<td>Lisbon, the Portuguese Erasmus city? Mis-match between representation in urban policies and international student experiences</td>
<td>03</td>
</tr>
<tr>
<td>Scientific journal paper</td>
<td>Confidence intervals for means and variances of nonnormal distributions</td>
<td>03</td>
</tr>
<tr>
<td>Scientific journal paper</td>
<td>Memories of (Un)Faire literacy policies in Southern Africa from the 1970s on: telling the (hi)story through life histories and photography of (dis)empowerment in Mozambique</td>
<td>03</td>
</tr>
<tr>
<td>Scientific journal paper</td>
<td>Internet of things and consumer engagement on retail: State-of-the-art and future directions</td>
<td>03</td>
</tr>
<tr>
<td>Scientific journal paper</td>
<td>The impacts of animal farming: A critical review of secondary and high school textbooks</td>
<td>03</td>
</tr>
</tbody>
</table>

Projects (317)

<table>
<thead>
<tr>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women architects in former Portuguese colonial Africa: gender and struggle for professional recognition (1953-1985)</td>
</tr>
<tr>
<td>Cooperation Project between the Youth Employment Observatory and the Portuguese Public Employment Services (IEFP)</td>
</tr>
<tr>
<td>Concepção e Implementação do Laboratório de Competências Transversais na Universidade Amilcar Cabral a partir do caso Iscte</td>
</tr>
<tr>
<td>Master’s Degree of Managing Digital Transformation in the Health Sector</td>
</tr>
<tr>
<td>Profissionalização artística e formação superior jazzística: a inserção profissional de jovens diplomados em Portugal</td>
</tr>
<tr>
<td>Trajetórias biográficas, percursos profissionais e inserção urbana de estudantes internacionais brasileiros em Lisboa</td>
</tr>
<tr>
<td>Flipping Learning Internationally in a Post Pandemic Era</td>
</tr>
<tr>
<td>11 TEIP - 2° Programa Territorialização de Políticas Educativas de Intervenção Prioritária</td>
</tr>
</tbody>
</table>
Applying Machine Learning
SDG Classification at Iscte
Timeline of our approach

June 2021 – September 2021
Testing/Training/Evaluating ML models

September 2021 – December 2021
Integration of ML model into information systems

July 2018 – December 2021
Faculty manually classifies outputs to SDGs

January 2022 – …
Faculty classifies outputs to SDGs with ML suggestions

euroCRIS Strategic Membership Meeting 2023
We are here!
### SDG Classification at Iscte

**Suggestions in our CRIS (Current Research Information System)**

<table>
<thead>
<tr>
<th>Type</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>The Sustainable Development Goals in higher education: a machine-learned approach</td>
</tr>
</tbody>
</table>

- **1. No Poverty**: Unselected
- **2. Zero Hunger**: Unselected
- **3. Good Health and Well-being**: Unselected
- **4. Quality Education**: Unselected
- **5. Gender Equality**: Unselected
- **7. Affordable and Clean Energy**: Unselected
- **8. Decent Work and Economic Growth**: Unselected
- **9. Industry, Innovation and Infrastructure**: Unselected
- **10. Reduced Inequalities**: Unselected
- **11. Sustainable Cities and Communities**: Unselected

*Considering the text content of this record, it is suggested that you choose this Sustainable Development Goal.*
Machine Learning Model for SDG Classification
Or how to teach a robot choose which SDG an output contributes to

SDG 1
SDG 2
“SDG 4” (with a certain level of confidence)

New Paper

Other Papers => SDGs
Machine Learning Model for SDG Classification

The actual machine learning training stage

Dataset → Clean/Prepare Data → Vectorize Data → Training Pipeline → ML Model
04 Evaluation
Evaluation of Machine Learning Algorithms

We tested the following algorithms with a dataset of 9665 annotated records:

- Gaussian Naive Bayes (GNB)
- Multinomial Naive Bayes (MNB)
- Linear Support Vector Classification (SVC)
- Logistic Regression (LR)

We used the following metrics:

- ACCURACY – ratio of correctly classified samples
- RECALL - ability to find all the positive samples
- PRECISION – ability to identify all the positive samples without accidentally marking too many negative samples as positive
- F-MEASURE - harmonic mean of the Precision and Recall
Comparison between Users Choices and Model Predictions

How much did the robot get right using the SVC-based Model?

Methodology

For each output, the user chose up to 3 SDGs.

Our Machine Learning model made the prediction for each combination of output/SDG, providing the corresponding score. The scores for each SDG were then sorted and only the top 3 were considered.

We then compared the two sets to see how much they matched.
Conclusions
Conclusions

Machine Learning Model is successful in suggesting SDGs that should be associated with outputs.

This work is not specific to SDGs. The same approach can be applied to other classification efforts.

We “automated” our users’ choices. That doesn’t mean they’re right. Expert review is needed.