

Pontifícia Universidade Católica de São Paulo
PUC-SP

Bárbara Melo - 3º semestre
Fabiana Campanari - 1º semestre
Gustavo Muriano - 1º semestre
Henrique Araujo - 3º semestre
Juliana Soares - 3º semestre
Luan Fabiano - 1º semestre
Pedro Gallego - 1º semestre
Rebecca Campos - 3º semestre

HackaPUCSP: Segunda edição

Ciência de Dados e Inteligência Artificial

São Paulo
2024

RESUMO

Durante o Hackathon de Ciência de Dados e Inteligência Artificial (IA) relacionado aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, realizamos uma análise dos dados disponibilizados pela universidade. O desafio proposto foi explorar as teses de diversos cursos da mesma, relacionando seus temas com os ODS, identificando áreas de destaque e possíveis ações visando o desenvolvimento da universidade enquanto organização comprometida com as políticas ESG colocando em prática habilidades técnicas, integração de novos alunos do curso e desenvolvimento tecnológico.

Palavras-chave: ODS; ONU; ESG; Hackathon; Desenvolvimento.

ABSTRACT

During the Data Science and Artificial Intelligence (AI) Hackathon related to the UN Sustainable Development Goals (SDGs), we carried out an analysis of the data made available by the university. The proposed challenge was to explore the theses of several courses within the same, relating their themes to the SDGs, identifying prominent areas and possible actions aimed at developing the university as an organization committed to ESG policies, putting technical skills into practice, integrating new students from course and technological development.

Keywords: SDGs; UN; ESG; Hackathon; Development.

LISTA DE ABREVIATURAS E SIGLAS

ODS	Objetivos de Desenvolvimento Sustentável
ONU	Organização das Nações Unidas
ESG	Environmental, Social e Governance
IA	Inteligência Artificial
PUC-SP	Pontifícia Universidade Católica de São Paulo
TEDE	Teses e Dissertações Eletrônicas

SUMÁRIO

1	INTRODUÇÃO	15
2	DESAFIO PROPOSTO	16
3	ODS E ESG.....	17
4	METODOLOGIA	18
4.1	Dados Utilizados	18
4.2	Código e etapas.....	19
4.3	Saída do código.....	23

1 INTRODUÇÃO

O objetivo do presente trabalho é realizar análises de dados e buscar previsões futuras a respeito dos compromissos públicos da PUC-SP relacionados as ODS para que se torne visível os méritos e desafios da instituição.

Para colocar em prática essas análises, é necessário monitorar quais as ODS se destacam na organização por meio de teses publicadas, recolhendo, armazenando e analisando dados e buscando justificativas para os resultados.

2 DESAFIO PROPOSTO

Foi disponibilizada uma base em TSV com um compilado de trabalhos acadêmicos já publicadas pela PUC-SP e cada grupo participante precisaria montar um código que lesse as teses e as-classificasse por ODS de forma a gerar Insights relevantes para a organização levantando hipóteses para as maiores ou menores ocorrências de temas para que assim, possam elaborar ações/ programas, afim de melhorar o cenário e incentivar o desenvolvimento sustentável, social e governamental da Universidade.

3 ODS E ESG

As ODS (Objetivos de Desenvolvimento Sustentável) são uma agenda global estabelecida pela Organização das Nações Unidas (ONU) para enfrentar os desafios mais urgentes do mundo. Para instituições, sejam elas públicas ou privadas, a importância das ODS é clara: oferecer um roteiro para direcionar esforços em direção a um desenvolvimento mais equitativo, sustentável e inclusivo.

Ao adotar as ODS, as instituições demonstram um compromisso com causas que ultrapassam suas fronteiras, engajando-se em uma agenda global de desenvolvimento sustentável. Isso não só melhora a imagem institucional, como também impulsiona a inovação e o crescimento sustentável e permitem avaliar e ajustar estratégias para garantir o compromisso contínuo com esses objetivos vitais para um futuro melhor.

Já as políticas ESG são diretrizes que orientam as práticas das organizações em relação aos aspectos ambientais, sociais e de governança. Elas têm uma importância crucial nas organizações, especialmente quando alinhadas com os Objetivos de Desenvolvimento Sustentável (ODS) da ONU.

A integração das políticas ESG com os ODS cria uma sinergia poderosa. As organizações passam a contribuir de forma direta para a realização desses objetivos globais, como a erradicação da pobreza, a ação contra as mudanças climáticas e a promoção da igualdade de gênero.

Assim, as políticas ESG se tornam não apenas uma necessidade ética, mas também uma estratégia inteligente para as organizações que buscam um papel significativo na construção de um mundo melhor, alinhado com os valores e objetivos das ODS.

4 METODOLOGIA

4.1 Dados Utilizados

[Link da tabela em TSV](#)

	A	B	C	D	E	F	G	H	I	J	K	L
1	datauridiomaseditoprogramadireito_acesodepartamentoarea_cnpqtitulotipotocinoleywordsabstractpalavras_chaveresumo											
2	2022-03-29https://rej.it evaluates the impac	it evaluates the inter technology and law	so the study is also g	it outlines the disrupti	but not without discu	da tecnologia e do c	assim sendo o estud	enfoça os efeitos dis	mas não sem discutir a	precafeidade na sur		
3	2019-08-27https://tec Política	SociedadeAcesso AI SP) - History Lyceu	SP) - History Secon	comprising the orgar from 1916	until 1951	when it underwent a	we enrol as the main	we mobilized the pui	as an analytical aid	w		
4	2010-05-18https://tec revelação e redes so	without exhausting it but only considering	the Map of Social N	disclosing does not r	living together	being always sociall	preserving a certain	each member of the	keeping their relator	reflecting on the reje	a	
5	2012-06-22https://tec we get to investigate	capital of Ceará in th	around the uses and	but also translated to	European especially the spaces of sociab	which recorded aspe	including those relat	shows up as research	important icons of m	but also significant v	a	
6	2006-08-25https://tec BA - História Educ	BA - História - 1870- BA - História - 1870	entre 1870/1890	implementado na pe	principalmente nas c	impôs novas demanc	principalmente a po	que deveria ser cons	moralizados	habilitados nos sabe	e	
7	2018-02-26https://tec the speech language specifically on the ci	it is justified by the d	along with the clinic	the instrument was a	on interaction situati	using AAC graphic a	and the possibility of	in the use of the AAC	easy to apply and lo	a intervenção fonoi	ju	
8	2006-10-11https://tec which in this case is	one proposes to revic	which from Philosoph	observed as languag	allows verifying its ac	the gift	that the paths of schi	from a more qualitat	and consequently	language-cognition	one seeks to compar	it
9	2006-11-09https://tec it details 203 articles	including sexual abu	rape	and the violation of	including online new witnesses	members of the Chui	and attorneys involv	preserving the anony	the analysis explores	the study points towa	e	
10	2011-06-10https://tec biológico	erégeno - um estudu	due the fact of being	a concept that dems	both physically and i	although in a slow m	obesity ends up prod	since each of them l	gathering them in a	as of the feeling of a	and the possibility th	a
11	2002-10-21https://tec a didactic sequence	who considers that a	that is	the corresponding lir	assessing their progr	com o auxílio de um	que considera impor	e leva em considera	iso é	a reta corresponden	de avaliar seus dese	e
12	2021-08-26https://rej Política	SociedadeAcesso AI	inspired by the Rom	being the Louis Brail	how they created the	in what context they	and which titles they	with the support of pl	reports	biographies	autobiographies	p
13	2006-08-03https://tec that had as focus of i	was considered to ve	in affirmative case	to describe which th	identifying possible f	developed in a daily	Paraná	had as participant of	they had been came	the two three times p	25 sessions of commi	ir
14	2018-02-05https://tec in the context of the	assuming that the co	and also contributed	reflecting on their fo	interviews were cond	making it possible to	drawing inspiration fi	1983). In both cases	the data were divide	based on the existi	we used Nôvoa (199	
15	2010-05-18https://tec and due to the succ	not to mention their	they deserve a top pi	but not al l long ficti	and have broadened	this study analyses th	Hiroyoshi Usui	Yasuhiko Iyoda	Hidekazu Nomura	and Inuhiko Yomota	denominados pelos j	d
16	2019-03-13https://tec bearing in mind the	the set of documents as well as a strategy	which adopted polic	tendo em vista a im	a base documental	t em que se discute a	assim como a estraté	que adotava políticas	que eram entendidas	como prejudiciais pa		
17	2007-05-11https://tec Brazil	understand dialog	in following the dialog	education toward hu	and with whom five r	associated with liber	though	show it is unassociat	instead	establish limits and	reaffirming its liberta	P
18	2011-06-18https://tec who is an engineerin	shows how the subje	under the author s su	started with the disc	by the instructors	of an engineering te	which was introduce	the Complex Thought	the instructors	based on pedagogic	critically reflected on th	
19	2008-04-09https://tec economical	social and cultural c	informed by the idea	progress and freedom	in a modernity-disc	most if all in Rio de	published between 1	an exploratory study	which included the j	in a euphoric period	in a discourse of mor	ir
20	2008-04-29https://tec LETRAS E ARTES-:I	in order to investigat	Brazil	through three main c	public	and out in t	based on genre and	in order to articulate	1987 and Dudley-Evi	1998)	the genre analysis m	1
21	2009-12-16https://tec according to the leg	that anticipated this	who assumed a dout	space	subject	and writing	their feeling about di	as well as their view	writing tasks and the	data shows their nee	conforme preconiza	e
22	20	município	Letras e ArtesCNPQ: LETRAS E ARTES-:I	Writing and Portugue	we defined the follow	the following specif	1996; SWALES	1990	2009; BHATIA	1993	2001; MILLER	2

Foram extraídos dados de milhares de trabalhos acadêmicos no TEDE escritos na última década e a partir disso, foi criado um dataset contendo títulos, resumos, palavras-chaves e diversos outros metadados potencialmente relevantes para a solução.

4.2 Código e etapas

```
#importar as bibliotecas que serão usadas
import pandas as pd
import os
import shutil
import re

# Caminho da pasta original
pasta_origem = '/content/drive/MyDrive/strings-de-busca-ods.zip
(Unzipped Files)'

# Criar uma nova pasta no Google Drive
nova_pasta = 'Nova_Pasta_Ods'
os.makedirs(nova_pasta, exist_ok=True)

# Lista de arquivos na pasta original
arquivos_na_pasta_origem = os.listdir(pasta_origem)

# Criar um dicionário de caminhos para os arquivos na nova pasta
caminhos = {}

for i, arquivo in enumerate(arquivos_na_pasta_origem, start=1):
    if arquivo.lower().endswith('.txt'): # Verificar se o arquivo é do
tipo txt
        nome_arquivo = f'ods{i:02d}.txt' # Formatação para garantir
dois dígitos (ex: 'ods01.txt')
        caminho_origem = os.path.join(pasta_origem, arquivo)
        caminho_destino = os.path.join(nova_pasta, nome_arquivo)

        # Copiar arquivo original para a nova pasta
        shutil.copy2(caminho_origem, caminho_destino) # Usar
shutil.copy2 para preservar metadados

        caminhos[f'ods{i:02d}'] = caminho_destino

# Iterar sobre o dicionário para carregar e processar os arquivos
for nome, caminho in caminhos.items():
    df = pd.read_csv(caminho, delimiter='\t')
    df = df.apply(lambda x: x.str.strip() if x.dtype == 'object' else
x)

    # Substituir palavras no DataFrame usando o dicionário alteracao
    alteracao = {
        "OU": "or",
        "E": "and",
        "NÃO": "not"
    }

    # Fazer alterações no arquivo, como traduzir para o inglês as
condições e tirar as barras e as chaves
    nome_coluna = df.columns[0] if len(df.columns) > 0 else None

    if nome_coluna is not None:
```

```

        df[nome_coluna] =
df[nome_coluna].str.replace('|'.join(alteracao.keys()), lambda x:
alteracao[x.group()], regex=True)
        df[nome_coluna] = df[nome_coluna].replace(r'[{}]', ' ',
regex=True)
        df[nome_coluna] = df[nome_coluna].replace(r'/', ' ',
regex=True)

        # Substituir o nome da coluna com base no nome do arquivo
        novo_nome_coluna = f'{nome}' # Adapte conforme necessário
        df.rename(columns={nome_coluna: novo_nome_coluna},
inplace=True)

        # Copiar o DataFrame modificado para um novo arquivo na pasta
de destino
        novo_caminho_destino = os.path.join(nova_pasta,
f'{nome}_modificado.txt')
        df.to_csv(novo_caminho_destino, sep='\t', index=False)

        print(f"Dataframe para {nome} após substituição:")
        print(df)

        # Se necessário, realizar o processamento adicional aqui (por
exemplo, remover valores vazios)
        df_sem_vazios = df.dropna(axis=1, how='all')
        treino = pd.read_csv('/content/train_df.tsv', sep='\t')
        teste = pd.read_csv('/content/test_df1.tsv', sep='\t')
        # Caminho da pasta com os arquivos das ODS modificadas
        pasta_ods_modificadas = 'Nova_Pasta_Ods'
        # Criar um dicionário para armazenar os DataFrames das ODS modificadas
        ods = {}
        # Iterar sobre os arquivos na pasta de ODS modificadas
        for arquivo_ods in os.listdir(pasta_ods_modificadas):
            if arquivo_ods.lower().endswith('_modificado.txt'):
                # Extrair o nome da ODS do nome do arquivo
                ods_nome = re.match(r'ods(\d+)_modificado\.txt',
arquivo_ods.lower())
                if ods_nome:
                    ods_nome = f'ods{ods_nome.group(1)}'
                    caminho_arquivo_ods = os.path.join(pasta_ods_modificadas,
arquivo_ods)
                    # Ler o arquivo como DataFrame e armazenar no dicionário
                    ods[ods_nome] = pd.read_csv(caminho_arquivo_ods,
delimiter='\t')
        # DataFrame chamado "teses" com as teses a serem classificadas
        teses = pd.read_csv('train_df.tsv', delimiter='\t')
        teses['ODS'] = '' # Adiciona uma coluna 'ODS' vazia para armazenar os
resultados
        colunas = teses.columns

```

```

# Dicionário para armazenar as teses classificadas para cada ODS
teses_classificadas = {key: [] for key in ods.keys()}
# Percorre cada tese e classifica de acordo com as chaves de busca de
cada ODS
for ods_nome, ods_df in ods.items():
    # Obter colunas específicas associadas à ODS (pode haver mais de
    uma coluna)
    colunas_chaves = [ods_nome] # Adicione mais colunas se necessário

    # Lista de listas de palavras-chave para a ODS (uma lista para cada
    coluna)
    palavras_chave_listas = [ods_df[col].dropna().tolist() for col in
    colunas_chaves]
    for index, tese_row in teses.iterrows():
        texto_tese = str(tese_row['palavras_chave']) # Supondo que o
        texto da tese está na coluna 'Texto'

        # Verifica se o texto da tese contém pelo menos uma palavra-
        chave de cada coluna da ODS
        if all(
            any(re.search(fr'\b{re.escape(palavra)}\b', texto_tese,
            flags=re.IGNORECASE) for palavra in palavras_chave)
            for palavras_chave in palavras_chave_listas
        ):
            teses_classificadas[ods_nome].append(index)
            teses.at[index, 'ODS'] = ods_nome # Adiciona a ODS
            correspondente à coluna 'ODS'
# Criar um objeto ExcelWriter para salvar as teses classificadas
excel_writer = pd.ExcelWriter('teses_classificadas.xlsx',
engine='xlsxwriter')

# Iterar sobre as ODS e salvar as teses classificadas em abas separadas
for ods_nome, indices_teses in teses_classificadas.items():
    print(f'Teses classificadas para {ods_nome}:')
    # Selecionar as teses classificadas para a ODS atual
    teses_ods = teses.iloc[indices_teses]
    # Salvar no ExcelWriter
    teses_ods.to_excel(excel_writer, sheet_name=ods_nome, index=False)
    print('\n')
# Fechar o ExcelWriter para salvar o arquivo
excel_writer.save()
# Percorre cada tese e classifica de acordo com as chaves de busca de
cada ODS
for ods_nome, ods_df in ods.items():
    # Obter colunas específicas associadas à ODS (pode haver mais de
    uma coluna)
    colunas_chaves = [ods_nome] # Adicione mais colunas se necessário

```

```

    # Lista de listas de palavras-chave para a ODS (uma lista para cada
coluna)
    palavras_chave_listas = [ods_df[col].dropna().tolist() for col in
colunas_chaves]
    print(palavras_chave_listas)

    for index, tese_row in teses.iterrows():
        texto_tese = str(tese_row['palavras_chave']) # Supondo que o
texto da tese está na coluna 'Texto'

        # Verifica se o texto da tese contém pelo menos uma palavra-
chave de cada coluna da ODS
        if all(
            any(re.search(fr'\b{re.escape(palavra)}\b', texto_tese,
flags=re.IGNORECASE) for palavra in palavras_chaves)
            for palavras_chaves in palavras_chave_listas
        ):
            teses_classificadas[ods_nome].append(index)
            teses.at[index, 'ODS'] = ods_nome # Adiciona a ODS
correspondente à coluna 'ODS'
# Percorre cada tese e classifica de acordo com as chaves de busca de
cada ODS
for ods_nome, ods_df in ods.items():
    # Obter colunas específicas associadas à ODS (pode haver mais de
uma coluna)
    colunas_chaves = [ods_nome] # Adicione mais colunas se necessário

    # Lista de listas de palavras-chave para a ODS (uma lista para cada
coluna)
    palavras_chave_listas = [ods_df[col].dropna().tolist() for col in
colunas_chaves]

    for index, tese_row in teses.iterrows():
        texto_tese = str(tese_row['palavras_chave']) # Supondo que o
texto da tese está na coluna 'palavras_chave'

        print(f'Tese #{index + 1} - ODS: {ods_nome}')
        print(f'Texto da Tese: {texto_tese}')

        # Verifica se o texto da tese contém pelo menos uma palavra-
chave de cada coluna da ODS
        if all(
            any(re.search(fr'\b{re.escape(palavra)}\b', texto_tese,
flags=re.IGNORECASE) for palavra in palavras_chaves)
            for palavras_chaves in palavras_chave_listas
        ):
            teses_classificadas[ods_nome].append(index)
            teses.at[index, 'ODS'] = ods_nome # Adiciona a ODS
correspondente à coluna 'ODS'

```

```

        print('Classificada!\n')
    else:
        print('Não classificada.\n')
# Criar um objeto ExcelWriter para salvar as teses classificadas
excel_writer = pd.ExcelWriter('teses_classificadas.xlsx',
engine='xlsxwriter')

# Dicionário para armazenar o número de teses classificadas para cada
ODS
num_teses_classificadas = {}

# Iterar sobre as ODS e salvar as teses classificadas em abas separadas
for ods_nome, indices_teses in teses_classificadas.items():
    print(f'Teses classificadas para {ods_nome}:')
    # Selecionar as teses classificadas para a ODS atual
    teses_ods = teses.iloc[indices_teses]
    # Salvar no ExcelWriter
    teses_ods.to_excel(excel_writer, sheet_name=ods_nome, index=False)
    num_teses_classificadas[ods_nome] = len(indices_teses) # Armazenar
o número de teses classificadas
    print(f'Número de teses classificadas para {ods_nome}:
{len(indices_teses)}\n')

# Imprimir a contagem total de teses classificadas para cada ODS ao
final do processo
print('\nResumo:')
for ods_nome, num_teses in num_teses_classificadas.items():
    print(f'{ods_nome}: {num_teses} teses classificadas')

# Fechar o ExcelWriter para salvar o arquivo
excel_writer.save()

```

4.3 Saída do código

A saída do código foi dividida nas três partes da saída, a primeira retorna as palavras-chaves relacionadas as ODS, a segunda classifica as teses em suas respectivas ODS de acordo com as palavras-chaves e na terceira e última, o código retorna com uma soma e resultado de quantas teses estão relacionadas a cada ODS. Com essa saída conseguimos identificar quais ODS possuem mais ou menos teses relacionadas a elas e associar possíveis razões para isso.

Retorna as palavras-chaves de cada ODS:

```
Dataframe para ods01 após substituição:
      ods01
0      (
1      inocência presumida
2      or falsa confissão
3      or conflito armado
4      or conflitos armados
..      ...
78     or liberdade de imprensa
79     or liberdade de expressão
80     )
81 and not ( doença or genética )
82     )

[83 rows x 1 columns]
Dataframe para ods02 após substituição:
      ods02
0      (
1      terrestre
2      or terra
3      or água doce
4      )
..      ...
71     or ecoturismo
72     or conservação baseada na comunidade
73 or conflito entre humanos e animais selvagens
74     )
```

Classifica as teses por ODS:

```
Tese #4837 - ODS: ods09
Texto da Tese: Textos dissertativo-argumentativos | SARES | Problema-Solução | Sistema de Avaliacao de Rendimento Escolar do Estado de Sao Paulo | Lingua portu
Não classificada.

Tese #4838 - ODS: ods09
Texto da Tese: Redator de textos técnicos | Ergologia | Análise do discurso de tradição francesa
Não classificada.

Tese #4839 - ODS: ods09
Texto da Tese: Persona | IMC | Obesidade | Roupa | Psicologia analítica
Não classificada.

Tese #4840 - ODS: ods09
Texto da Tese: Culpa | Responsabilidade | Nazismo | Alemanha | Arendt | Jaspers
Não classificada.

Tese #4841 - ODS: ods09
Texto da Tese: Formação permanente de professores | Saberes docentes | Pesquisa intervenção colaborativa | Professores -- Formacao profissional | Educacao cont
Não classificada.

Tese #4842 - ODS: ods09
Texto da Tese: Temas Transversais | Professores de ciência - Formação profissional | Ensino fundamental - Currículos - Brasil
Não classificada.

Tese #4843 - ODS: ods09
Texto da Tese: ritos | templo | sacerdote | Cultos afro-brasileiros | Relacoes sociais | Ritual | Ciencias Sociais
Não classificada.
```

Soma a quantidade de teses relacionadas a cada ods:

```
Resumo:
ods08: 36 teses classificadas
ods07: 36 teses classificadas
ods12: 18 teses classificadas
ods13: 36 teses classificadas
ods16: 36 teses classificadas
ods01: 36 teses classificadas
ods03: 66 teses classificadas
ods11: 15 teses classificadas
ods14: 360 teses classificadas
ods02: 66 teses classificadas
ods15: 15 teses classificadas
ods06: 513 teses classificadas
ods10: 60 teses classificadas
ods04: 48 teses classificadas
ods05: 36 teses classificadas
ods09: 36 teses classificadas
```