



Projeto de Formatura – 2023 – Press Release
PCS - Departamento de Engenharia de Computação e
Sistemas Digitais

Tema: Método de classificação de empresas sustentáveis desenvolvido com NLP

Estudante da Poli-USP desenvolve método de classificação de de empresas sustentáveis com NLP e interface para o uso do modelo.

São Paulo, 28 de novembro

Nos dias 19 e 20 de Dezembro de 2023, os estudantes Marcelo Dias e Rafael Rahal do curso de Engenharia da Computação, da Escola Politécnica da USP, apresentarão seu trabalho de conclusão de curso, desenvolvido sob orientação do Professor Doutor Pedro Corrêa, à banca de professores avaliadores.

O projeto se trata de um modelo de classificação de empresas como sustentáveis ou não com base em sua descrição de texto. Usando uma arquitetura de modelo chamada Transformer, o grupo utiliza de um modelo pré-treinado chamado BERT, que consegue comprimir e compreender sentenças.

Para disponibilizar o uso do modelo, a equipe criou também uma interface de comunicação com o sistema. Tal interface permite o usuário colocar a descrição de sua empresa ou até utilizar de exemplos já criados para observar o comportamento do modelo. Não só, possui algumas visualizações do dataset usado, trazendo ainda mais informação sobre sustentabilidade ao usuário.

O dataset usado foi obtido através da página Crunchbase, um site de aporte de investimentos que contém mais de 23 mil empresas cadastradas. Além disso, possui também diversas categorias que a empresa pode selecionar, sendo uma delas de sustentabilidade. Esta foi usada como o objetivo para treinar o modelo.

As tecnologias usadas para desenvolver o sistema foram React, para a interface web, Python para o servidor que conecta com a interface e Python também para a implementação do modelo. Não só, foram usadas bibliotecas como HuggingFace, TensorFlow e Sklearn para conseguir desenvolver e treinar a máquina.

Integrantes: Rafael Rosa Rahal
Marcelo Dias de Oliveira Fernandes

Professor(a) Orientador(a): Prof. Dr. Pedro Luiz Pizzigatti
Corrêa
