**Resumos para os posts do C2D sobre tecnologias de Machine Learning**

(Tentar se limitar a 2, no máximo 3 parágrafos → até 12 linhas.)

1. BERTopic: (Não existe paper sobre)

BERT [1] é atualmente uma das arquiteturas de rede neural mais utilizadas e, recorrentemente citada como o atual estado da arte em diversas tarefas de processamento de linguagem natural (PLN). O principal motivo é pela abstração de contexto e entendimento de linguagem que os blocos de Encoders [2] possibilitam associado com a facilidade de *fine tuning* da rede em tarefas específicas como tradução automática, sumarização de textos ou análise de sentimentos, por exemplo.

Uma dessas tarefas é a de *topic modeling*, que visa entender qual o principal assunto em uma coletânea de documentos a partir dos tópicos mais recorrentes. Para isso foi desenvolvido o BERTopic [3] que utiliza os Transformers de sentença para criar vetores para os documentos de entrada e gerando 2 saídas possíveis: tópicos que englobam um conjunto de palavras sobre um mesmo assunto e probabilidades que mostram o quanto um documento está relacionado a um possível tópico.

[1] [BERT - Jacob Devlin (2018)](https://arxiv.org/pdf/1810.04805.pdf)

[2] [Attention Is All You Need - Ashish Vaswani (2017)](https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)

[3] [BERTopic - Maarten Grootendorst (2020)](https://github.com/MaartenGr/BERTopic)

1. Sumarização Automática

Com a quantidade de informação textual que temos contato diariamente a Sumarização Automática é uma técnica de processamento de linguagem natural (PLN) cada vez mais relevante que visa tornar um grande conjunto de textos em um documento conciso e que mantenha as informações chave do documento original, essa tarefa é especialmente desafiadora devido ao grau de subjetividade e interpretação de mundo que são necessários para a construção de um bom resumo.

Os métodos de Sumarização Automática tem duas principais abordagens: extração e abstração. O primeiro consiste em identificar e retirar as principais passagens dos textos, filtrando redundâncias e justapondo essas frases em um texto enxuto. O segundo método é mais parecido com a abordagem humana, ou seja, a partir de um conjunto de informações recriar um novo conjunto de texto a baseando-se em refraseamento, entendimento e sinônimos para reduzir o volume textual.

1. Análise de Sentimentos