EXPLORANDO AS FUNÇÕES DE ROTULAGEM DO SNORKEL PARA TEXTOS DA ÁREA DE FARMACOVIGILÂNCIA

Nos últimos anos o avanço em aprendizado de máquina tem se dado em parte a *datasets* rotulados. Estes *datasets,* porém, ainda são extremamente custosos de se produzir devido seu volume e pela necessidade do entendimento do domínio. Este estudo buscou entender a ferramenta Snorkel, capaz de automatizar este processo de rotulagem dos dados, permitindo o gerenciamento de dados de treinamento para uso nos métodos de aprendizado de máquina. A ferramenta oferece suporte para o usuário rotular, construir e gerenciar de maneira programática os dados de treinamento. O projeto também analisou a viabilidade da ferramenta para o uso na língua portuguesa usando textos a serem rotulados da área de farmacovigilância. A metodologia deste estudo inicialmente envolveu o entendimento, configuração, e execução da ferramenta Snorkel por meio de seus tutoriais e *datasets* disponibilizados. Em um segundo momento, as variações de funções de rotulagem foram exploradas para a execução da ferramenta com foco no domínio de estudo, a farmacovigilância. Para a execução foram utilizados *datasets* da área de farmacovigilância para a busca em *tweets* de relatos de eventos adversos causados por medicações. Como resultados, verificou-se que por mais que haja problemas com a ferramenta devido às suas diferentes versões, sua instalação e execução ainda são possíveis, porém com certos ajustes. Averiguou-se também que para o uso da língua portuguesa é necessária a criação ou adaptação de *datasets* para que os mesmos estejam somente em português, assim como o desenvolvimento de funções de rotulagem que se adaptem aos dados do domínio. Tais adaptações foram implementadas utilizando a linguagem de programação Python. A partir desses resultados, pode-se considerar que o uso da ferramenta é viável para ser utilizada na língua portuguesa e que seu uso pode ser aplicado ao domínio de farmacovigilância com registros em linguagem natural, podendo apoiar a área no Brasil.