Medicrawl: An Ml Based System to Automate the Search of Illegal Products in Ecommerce

Nicolas Perez Gonzalez

`nicolas.perezgonzalez@swissmedic.ch`

Abstract

Swissmedic is the National Authorisation and supervisory authority for drugs and medical products. It ensures high-quality, safe and effective medical products. As part of its mandate, Swissmedic regularly oversees online markets on the search for illegal medicinal products. This search is time consuming as well as resource intensive. In order to optimize this workflow, MediCrawl was developed in the innovation team within Swissmedic. MediCrawl is a platform developed in the Cloud that allows a centralized analysis of products of interest found online. These products are collected via crawling and are later presented to the Swissmedic specialists. Due to the large number of incoming products of interest, a series of filters have been developed including a Machine Learning based filter based on the text description of these. With a dataset of roughly 60,000 cases, we trained multiple classifiers based on BERT embeddings, which identifie the binary signal and labels if the case is relevant for Swissmedic. In development, our classifier identified relevant products with a precision of 0.85 and recall of 0.92. 3 months later, with new unseen data, precision dropped to 0.61 while the relevant score of recall remain high at 0.94. After two years of development, our results with MediCrawl are promising and a more permanent implementation in Swissmedic is underway as well as an international collaboration for this tool. Further research will be focused on improving scores with new methodologies as well as the benefits of moving training into a MLOps pipeline.