Loblaw Digital: Improving Product Categorization Using NLP Techniques

To organize their online product catalogue, Loblaws manually constructs hierarchical product taxonomies. These classifications, along with historical product purchases, are used to improve the search results on their online shopping platform. This strategy for improving the search capabilities has two weaknesses when used on new products in catalogue - the item has to be manually placed into a product categorization that is extremely large, and items without a purchase history run into the cold start problem because they have no purchases to be compared to. Our team seeks to address both of these problems using machine learning. Using the current product classifications as our labels, we will use product metadata (product names and descriptions) to classify each item. We will use text encoding to create meaningful training data for our own algorithm, that will also help Loblaw Digital with the cold start for new products. Our team has created classification models using sklearn, with up to 87% accuracy using basic NLP techniques on the metadata provided. We have also explored more advanced vector embedding techniques on the dataset which have given promising results. We will continue to improve our model performance as our work moves on.