Product Deduplication

Identified similar products for e-commerce websites and built a system that can detect and de-duplicate them. Using distributed data systems, pre trained word and image embeddings coupled with machine learning/neural networks algorithms like logistic regression and approximate KNN. Achieved an accuracy of 0.73 on the best model.