

Website Search - Keyword Optimization

Problem Statement & Exploratory Data Analysis

Objective : Providing contextual search results by analyzing historical search patterns to identify consumer interest and keywords and associate products with high value keywords.

Business Benefits : Improve customer shopping experience and better search conversion

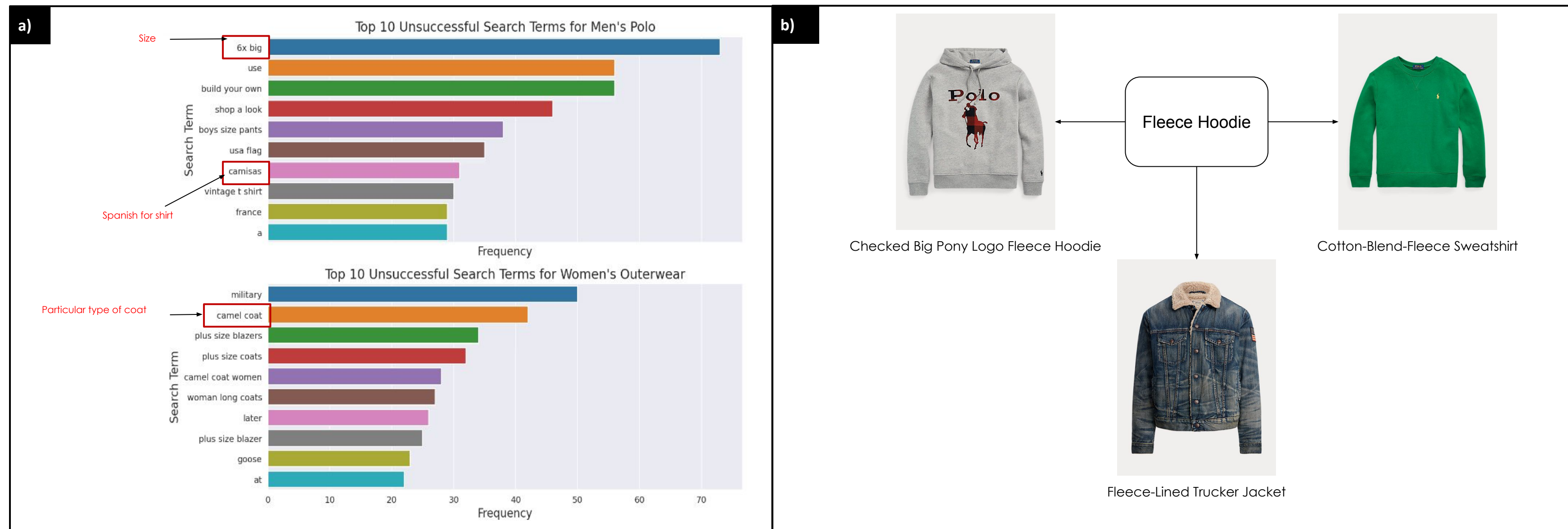


Figure 1. (a) Unsuccessful searches that do not result in any products (b) Products bought using the search term "Fleece Hoodie" - evident semantic relation

Approach 1: Search-Product Embedding Similarity

The motivation for model 1 was to build a product recommendation model that is robust to unseen data (new search terms and products). This model uses semantic similarity between embeddings of product metadata and search terms to make recommendations.

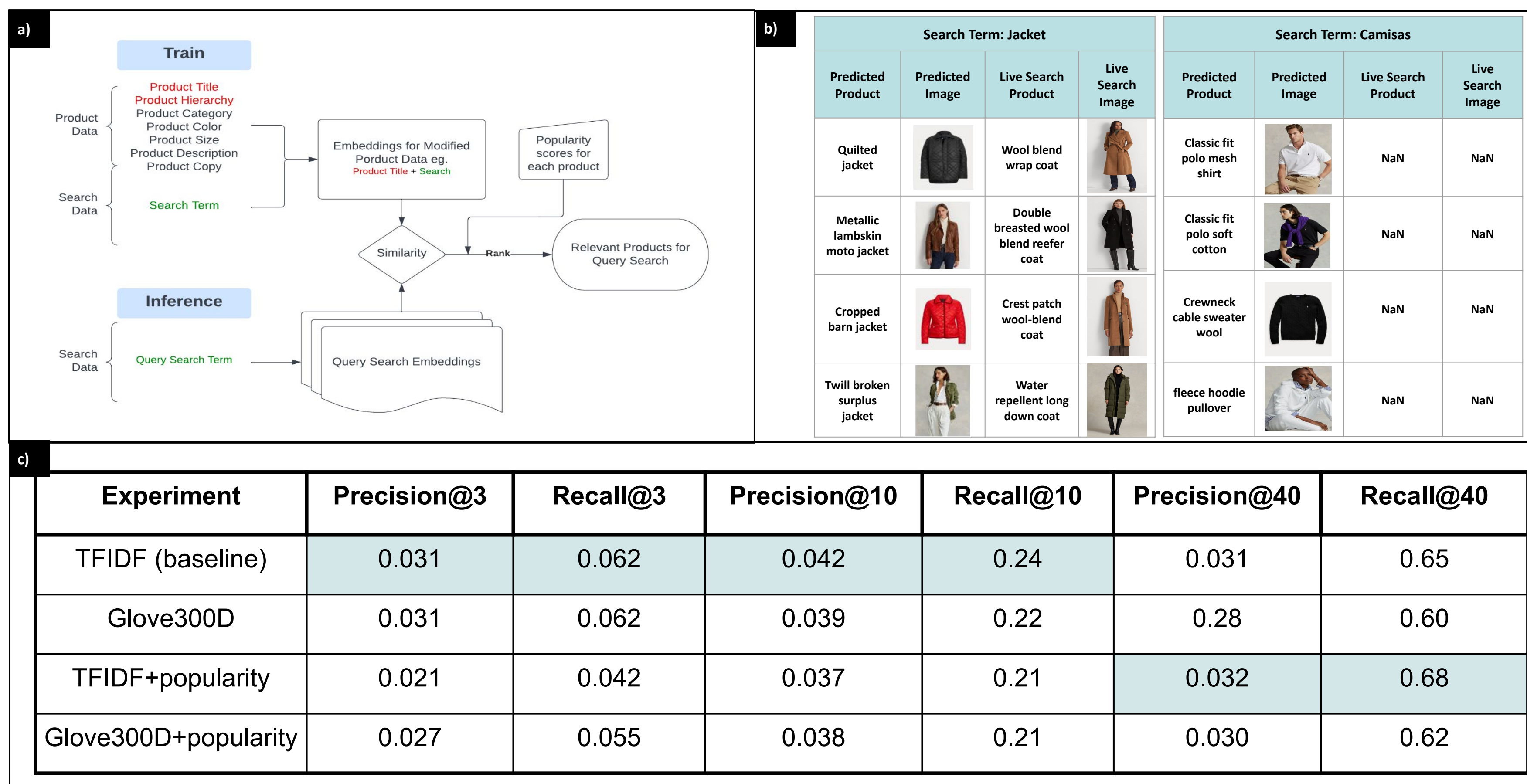


Figure 2. (a) Method flow (b) Example search results using TF-IDF+popularity (c) Precision and recall of predictions at various ranks

Approach 2: Search-Product Pair Classification

The motivation behind model 2 lies in the robustness to new products without the need for retraining and also to learn semantic similarity and popularity from the data.

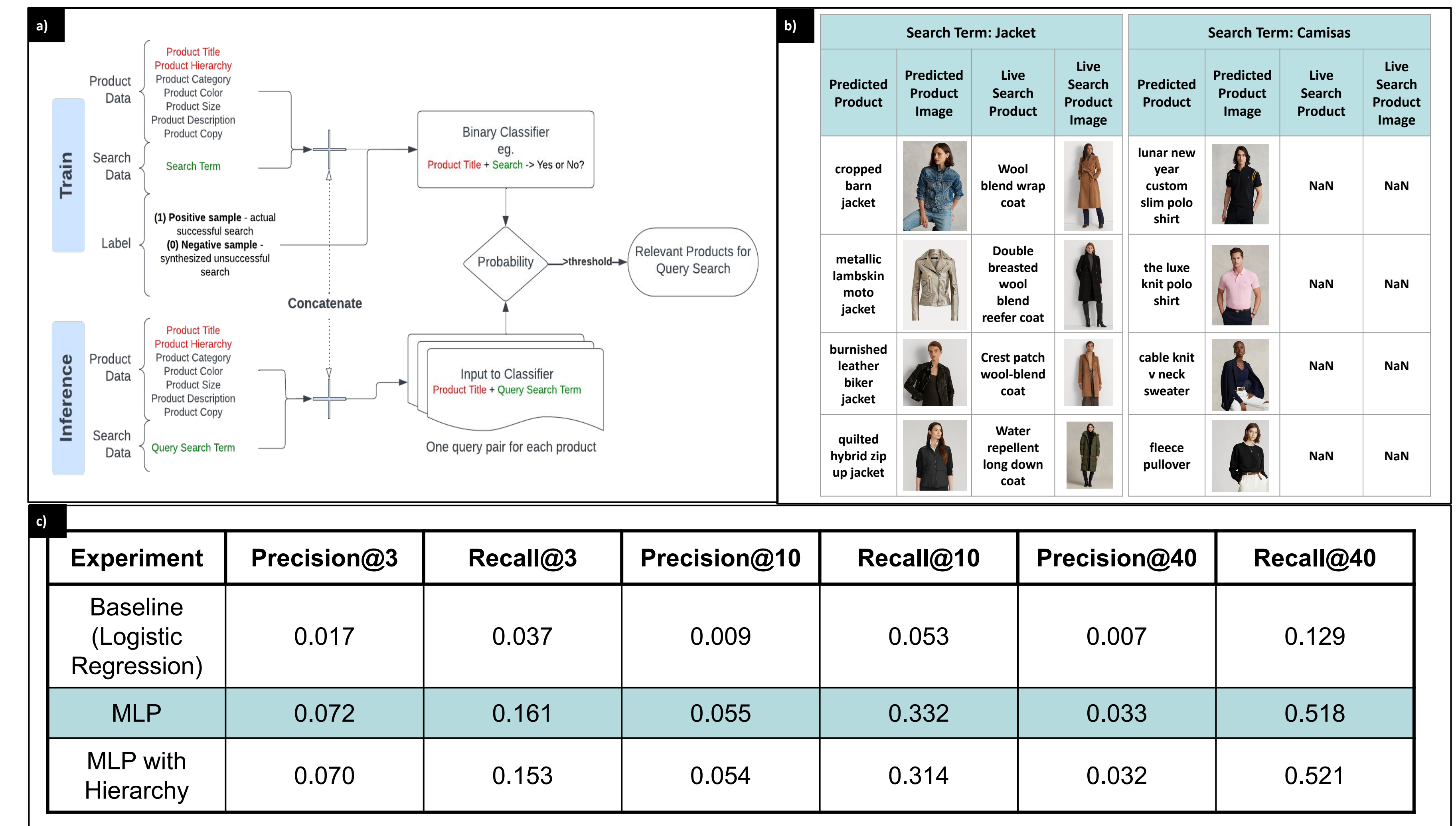


Figure 3. (a) Method flow (b) Example search results using MLPClassifier (c) Precision and recall of predictions at various ranks

Conclusions and Future Work

Our best performing model was the MLP without Hierarchy - especially among the top 3 and 10 products for each search. At k=40, TF-IDF similarity with popularity also performs well. We plan to investigate how the advantages of TF-IDF and popularity can be incorporated into the MLP model. In addition, we would want to experiment with BERT embeddings in place of GloVe.

Acknowledgments

We are grateful to the Ralph Lauren team for their mentorship and guidance throughout the project. We would also like to thank Prof. Kelleher and the rest of the staff for their support.

References

- [1] Xiong, Y., Chang, W. C., Hsieh, C. J., Yu, H. F., & Dhillon, I. (2021). Extreme Zero-Shot Learning for Extreme Text Classification. arXiv preprint arXiv:2112.08652.
- [2] Sorokina, D. and Cantu-Paz, E. (2016). Amazon search: The joy of ranking products. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 459-460).