AI for Marketing: Final Report

Paula Jeniffer dos Santos Viriato[1][0000−0003−0900−1686]

Institute of Computing, University of Campinas, Campinas - SP, Brazil
p234831@dac.unicamp.br

1 Project State in 2023 May - NIHER for Marketing: Non-invasive and Highly Explainable Recommendations for Marketing

As shown in Figure 1, this project aims to overcome the current state of digital marketing by promoting highly explainable and non-intrusive recommendations. We will see in the next section that high explainability, linked to respect for users' privacy, is critical to ensuring ethics in artificial intelligence. Next, we will see the contents present in the bibliography essential for the fulfillment of the objectives of this project.

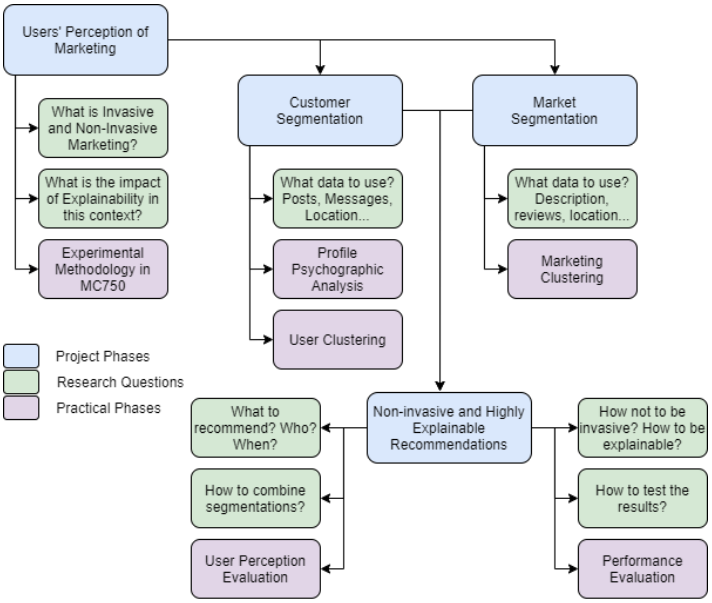

Fig. 1: Non-invasive and Highly Explainable Recommendations - Project Overview

1.1 Bibliographic Review

Currently, digital marketing offers several advantages to the market, such as establishing a high level of interactivity with the consumer, facilitating the

segmentation and targeting of customers, generating convenience for the consumer, and requiring less investment. However, there are still many barriers to be overcome. [16]

The online reputation of companies can be destroyed by negative feedback, as brands are visible and accessible to all users on the internet. There is a lack of consumer confidence due to the tracking and collection of user data for the technological development of personalized recommenders, generating a conflict related to the privacy of the data. Digital marketing is highly dependent on [7, 8] techniques and technologies, and many companies do not make good use of the available tools and applications. [16] In this way, making digital marketing more transparent and comfortable for users is also necessary for companies.

Explainable Artificial Intelligence (XAI) Artificial intelligence algorithms are applied in various marketing-oriented applications, such as personalizing recommendations, both for products and content, and such content ranges from news to knowledge sources and social media posts. Artificial intelligence also optimizes the financial returns gained from targeted marketing, using data mining of online consumer behavior (Customer Behavior) in conjunction with ad segmentation. However, many machine learning algorithms have very obscure decisions and behaviors, especially when we are talking about deep neural networks (Deep Learning) that are difficult to explore. Such behavior undermines customer confidence with possible algorithmic biases, leading to large-scale discrimination. [13]

Explainable Artificial Intelligence (XAI) is a class of systems that provide visibility into decision-making, making it fundamental for marketing research. Using personal information in personalized recommendations and mobile advertising raises privacy concerns for users. XAI is an important area of research that influences the calculation of individuals' privacy, being able to meet such demand in algorithms aimed at Marketing. [13]

Highly Explainable Recommender Systems Modern recommender systems make use of explicit and implicit features for personalization. However, such recommendation models suffer from a lack of explainability, and many customers are dissatisfied with this lack of clarity. Models based on regressions and trees are more explainable, but there are alternatives with lesser explainability using models based on matrix factorization [6, 9], collaborative neural filtering [4, 5], generative adversarial networks [18, 20], and graph-based models [3], which have grown a lot in the last decade. [19]

Adding contextualization features is seen as a solution for greater explainability of recommendation-focused machine learning models, and the most used features are locations, age, gender (metadata), images, titles, and descriptions of previously accessed items. Sentiment analysis based on comments about products is another resource currently widely used. Aiming to increase the recommendation's explainability, architectures with different levels of transparency are being studied: whitebox models with attention to features, greybox models with adversarial training, and blackbox models with counterfactual increases. [19]

Custumer Segmentation Customer segmentation is used as an input to customize services based on current customer conditions, and is performed by processing customer data. [14] Market segmentation is defined as the division of a market according to the characteristics and needs of its customers, culminating in different marketing efforts [1]. On the other hand, customer segmentation takes the focus away from the market and puts the focus on the customer. The objective is to group the customer based on different schemes: value, behavior, lifestyle, life cycle, and activity. The client can be segmented by supervised and clustering methods, and by clustering, the clients are grouped more resistant to outliers, and [11] anomalies. Customer groupings can occur both descriptively and predictively. [2]

Market segmentation is one of the most important market concepts, but customer segmentation is an effective way of marketing personalization and assertiveness regarding customer needs. [2]

Users' Perception of Marketing In the article by Puranik and Bansal [12], exploratory research was carried out on the perception of customers about online shopping. This study used a sample of 100 respondents from a city in India. As a result of a factor analysis, seven key factors emerged in the users' perception of online shopping. Ordered from the factor with the highest factorial load to the factor with the lowest factorial load, we have Relevant Information, Reliability, Previous Experience, Instant Review, Product Delivery, Transparency, and Seller Image.

In this way, it is possible to adapt the methodology of Viriato *et al.* [17], on the evaluation of humanization in chatbots to adapt the factors observed in Puranik and Bansal [12] for each case of placement of marketing. The association of the study by Puranik and Bansal [12] with Viriato *et al.* [17] makes it possible to study the generation of a user-centered marketing platform and its comfort.

1.2 Experimental Methodology

Our objective is twofold: to make recommendations based on user text, which in our experimental case comes from forums, and recommend marketing content posting locations based on their textual content. Figure 2 shows that data extraction occurs similarly for texts produced by users (media on Reddit) and textual content of products on the Amazon website.

We use the search system of both platforms to find similar thematic content and, thus, labels. Contents on the following topics were selected, based on Reddit's own content categorizations: game, sport, business, economic, finance, television, celebrity, animal, pet, anime, art, car, motor vehicle, motorcycle, crafts, diy, culture, race, ethnicity, fashion, food, drink, history, hobby, law, learn, education, military, movie, film, music, place, podcast, stream, politic, programming, reading, writting, literature, religion, spirituality, technology, and travel.

There is a similarity between the two proposed recommenders. Given a specific topic, products on this topic should be recommended to users who write about

this topic in forums (Product Recommendation for Customers). On the other hand, forums on the same topic as the product should be recommended for posting in the respective digital marketing (Location Recommendation for Marketing). Because of this characteristic among recommenders, we are researching and verifying the usability of Siamese networks for this case.

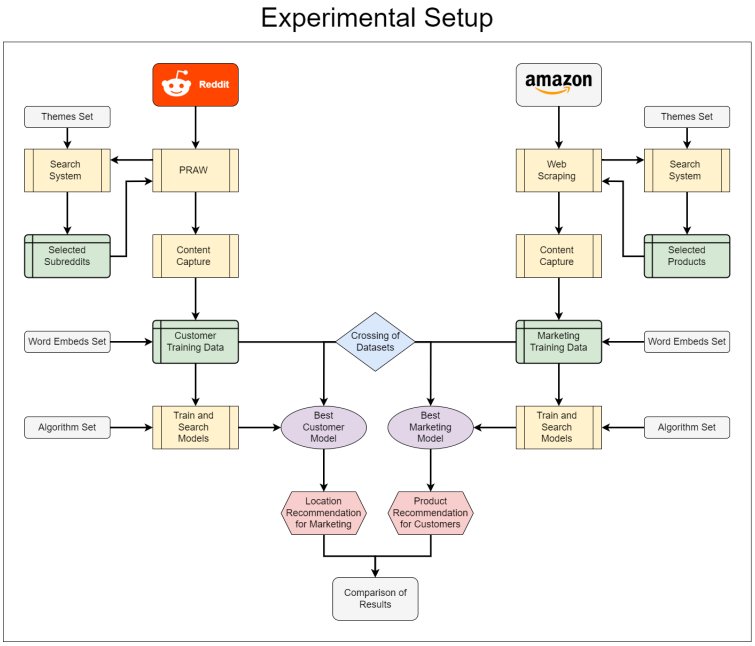

Fig. 2: Experimental Methodology Setup

1.3 Datasets

Dataset from Reddit An alternative thought about a dataset in which it is possible to verify more precisely the interests of the users is the extraction of data from Reddit forums. Reddit is the self-proclaimed "front page of the internet" [15], being the world's largest online discussion forum [10]. It is organized into topics such as music, movies, gardening, computing, and more. The extraction of data from these forums focused on specific themes can help in a stronger recommendation for users.

1.4 Preliminary Results

Extraction of posts and comments from Reddit according to theme. The first topics worked on in Portuguese were: animals, beauty, science, construction, cooking, design, sports, photography, games, gardening, books, music, technology,

vehicles, travel, and videos. In this first application of the Reddit data extractor, present in the following link, we obtained 17.386 posts. The merging of posts from all themes into a single dataset and some statistics can be acquired by the check_data code. The number of posts extracted for each of the themes is shown in Table 1.

Table 1: Number of Posts Acquired by Subreddit

animals: 2403	cooking: 590	games: 205	technology: 1080
beauty: 180	drawing: 866	gardening: 200	vehicles: 1329
science: 2068	sports: 2101	books: 1396	trips: 293
build: 10	photography: 987	music: 1997	videos: 1681

Figure 3 shows how much each theme affects the overall design of the dataset so far, by means of a pie chart. Here we have already managed to analyze that some themes participate more in the composition of the Dataset than others, such as animals, science, books, music, videos and sports. This could be happening for three reasons. The first one due to a real trend on Reddit for certain subjects. The second justification would be an imbalance of Reddit as a whole for Reddit in Portuguese. The third justification would be that not enough data has been mined yet. For the second case, an alternative would be to translate subreddits from other topics, using libraries already implemented for this purpose; and for the third case, the alternative would be to continue mining more and more data.

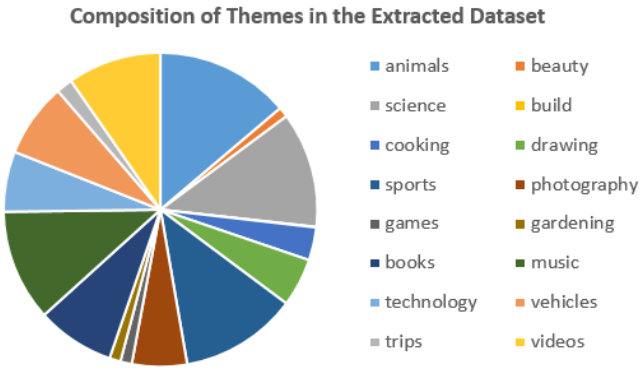

Fig. 3: Composition of Themes in the Extracted Dataset

References

1. Armstrong, G., Adam, S., Denize, S., Kotler, P.: Principles of marketing. Pearson Australia (2014)
2. Cooil, B., Aksoy, L., Keiningham, T.L.: Approaches to customer segmentation. Journal of Relationship Marketing **6**(3-4), 9–39 (2008)
3. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: Lightgcn: Simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. pp. 639–648 (2020)
4. He, X., He, Z., Du, X., Chua, T.S.: Adversarial personalized ranking for recommendation. In: The 41st International ACM SIGIR conference on research & development in information retrieval. pp. 355–364 (2018)
5. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S.: Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide web. pp. 173–182 (2017)
6. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
7. Krasnov, S., Sergeev, S., Mukhanova, N., Grushkin, A.: Methodical forming business competencies for private label. In: 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). pp. 553–558. IEEE (2017)
8. Lichy, J., Kachour, M., Khvatova, T.: Big data is watching you: opportunities and challenges from the perspective of young adult consumers in russia. Journal of Marketing Management **33**(9-10), 719–741 (2017)
9. Ma, H., Yang, H., Lyu, M.R., King, I.: Sorec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM conference on Information and knowledge management. pp. 931–940 (2008)
10. Medvedev, A.N., Lambiotte, R., Delvenne, J.C.: The anatomy of reddit: An overview of academic research. Dynamics On and Of Complex Networks III: Machine Learning and Statistical Physics Approaches 10 pp. 183–204 (2019)
11. Punj, G.N., Stewart, D.W.: An interaction framework of consumer decision making. Journal of Consumer Research **10**(2), 181–196 (1983)
12. Puranik, R., Bansal, A.: A study of internet users' perception towards e-shopping. Pacific Business Review International **6**(9), 37–44 (2014)
13. Rai, A.: Explainable ai: From black box to glass box. Journal of the Academy of Marketing Science **48**, 137–141 (2020)
14. Sari, J.N., Nugroho, L.E., Ferdiana, R., Santosa, P.I.: Review on customer segmentation technique on ecommerce. Advanced Science Letters **22**(10), 3018–3022 (2016)
15. Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., Strohmaier, M.: Evolution of reddit: from the front page of the internet to a self-referential community? In: Proceedings of the 23rd international conference on world wide web. pp. 517–522 (2014)
16. Veleva, S., Tsvetanova, A.: Characteristics of the digital marketing advantages and disadvantages. In: IOP Conference Series: Materials Science and Engineering. vol. 940, p. 012065. IOP Publishing (2020)
17. Viriato, P.J.d.S., Souza, R.R.d., Villas, L.A., dos Reis, J.C.: A framework for humanization evaluation in chatbots. In: HCI International 2023: 25th International Conference on Human-Computer Interaction. Computer and Information Science (CCIS), Springer, Copenhagen, Denmark (2023)

18. Wang, J., Yu, L., Zhang, W., Gong, Y., Xu, Y., Wang, B., Zhang, P., Zhang, D.: Irgan: A minimax game for unifying generative and discriminative information retrieval models. In: Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval. pp. 515–524 (2017)
19. Zhou, Y., Wang, H., He, J., Wang, H.: From intrinsic to counterfactual: On the explainability of contextualized recommender systems. arXiv preprint arXiv:2110.14844 (2021)
20. Zhou, Y., Xu, J., Wu, J., Taghavi, Z., Korpeoglu, E., Achan, K., He, J.: Pure: Positive-unlabeled recommendation with generative adversarial network. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 2409–2419 (2021)