Literature survey

Team	Abhiram Grandhi Nikil Jonnada Sadana Ragoor

Title: AI-Powered Customer Personalization Platform

Scope

This project aims to design, develop, and implement an AI-Powered Customer Personalization Platform that leverages machine learning, customer data analytics, and real-time insights to enhance customer experiences in an e-commerce or marketing context.

The primary goal is to develop, implement, and launch a platform that utilizes artificial intelligence (AI) and machine learning (ML) technologies to improve customer experiences.

The platform will cater to businesses operating in e-commerce or marketing domains, offering personalized services and recommendations to their customers.

The project will focus on three key features: dynamic product recommendations, behavioral segmentation, and real-time customer insights.

Key Features:

1. **Dynamic Product Recommendations:**
    - AI algorithms will analyze user behavior, including browsing and purchase history.
    - Based on this analysis, the platform will suggest personalized product recommendations to users in real time.

- This feature aims to enhance product discovery and increase user engagement and conversions.

2. **Behavioral Segmentation:**
   - The platform will employ AI techniques to group customers into distinct segments based on their behavior, preferences, and interactions with the platform.
   - These segments will facilitate targeted marketing campaigns, allowing businesses to tailor their messaging and promotions for maximum effectiveness.
3. **Real-time Customer Insights:**
   - The platform will provide real-time analytics and insights regarding customer preferences, behaviors, and trends.
   - These insights will enable businesses to make data-driven decisions and adapt their strategies in real time, improving customer satisfaction and retention.

Search Strategy

To conduct a thorough literature review for our project titled "AI-Powered Customer Personalization Platform," we employed a well-structured search strategy to identify relevant sources from various databases, libraries, and online resources. The strategy involved the following key steps:

1. Selection of Relevant Databases and Resources:

We identified and utilized several reputable academic databases and resources, including

- Google Scholar
- PubMed (for healthcare-related personalization)
- IEEE Xplore (for technical and engineering aspects)
- ACM Digital Library (for computer science and technology)
- JSTOR (for humanities and social sciences)

- Business Source Premier (for business and marketing research)
- Kaggle for datasets

2. **Compilation of Keywords and Search Terms:**

   We developed a comprehensive list of keywords and search terms related to our research topic, ensuring the inclusion of synonyms, variations, and relevant phrases. These keywords included, but were not limited to,

   - "AI-Powered Customer Personalization,"
   - "Dynamic Product Recommendations,"
   - "Behavioral Segmentation,"
   - "Real-time Customer Insights."
   -

3. **Dataset Acquisition from Kaggle:**

   To kickstart our project, we opted to leverage Kaggle as a valuable resource for acquiring initial datasets pertinent to our research on the "AI-Powered Customer Personalization Platform." Kaggle is a renowned platform in the data science and machine learning community, offering a diverse collection of high-quality datasets contributed by data enthusiasts, researchers, and organizations worldwide. The platform not only provides access to a wide array of datasets but also fosters a collaborative environment where data scientists and analysts can share, explore, and collaborate on data-related projects. For our project, we carefully selected datasets from Kaggle that align with our research objectives, enabling us to train and test our machine learning models, conduct data analytics, and derive insights into customer behavior and preferences. These datasets will serve as a foundation for building and fine-tuning our personalization algorithms and analytical tools.

   example datasets:

> 💡 https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store

This file contains behavior data for 7 months (from October 2019 to April 2020) from a large multi-category online store.

Each row in the file represents an event. All events are related to products and users. Each event is like many-to-many relation between products and users.

Data collected by Open CDP project. Feel free to use open source customer data platform.

More datasets

Checkout another datasets:

1. https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-electronics-store
2. https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop
3. https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-jewelry-store
4. https://www.kaggle.com/mkechinov/ecommerce-events-history-in-electronics-store
5. [NEW] https://www.kaggle.com/datasets/mkechinov/direct-messaging

Selection Criteria

Selection Criteria for Inclusion:

1. **Relevance to Research Objectives:**

- The primary criterion for inclusion is the relevance of the source to our research objectives. Sources must directly contribute to our understanding of AI-driven customer personalization, dynamic product recommendations, behavioral segmentation, and real-time customer insights.

2. **Publication Date:**
    - Given the rapidly evolving nature of AI and machine learning, we prioritize recent sources published within the last 2-4 years. This ensures that our review incorporates the latest advancements and insights in the field.
3. **Credibility and Source Type:**
    - Peer-reviewed journal articles, conference papers, and reputable academic publications hold priority due to their rigorous review processes. Sources from recognized scholars, institutions, and industry experts are considered more credible.
4. **Methodology and Empirical Studies:**
    - Empirical studies that employ sound research methodologies, including experiments, surveys, case studies, and data-driven analyses, are preferred. These studies provide concrete evidence and insights into the effectiveness of AI-powered personalization techniques.
5. **Theoretical Frameworks and Conceptual Papers:**
    - Theoretical and conceptual papers that offer frameworks, models, or theoretical perspectives relevant to AI-driven customer personalization are included. These sources help us establish a theoretical foundation for our project.
6. **Cross-Disciplinary Insights:**
    - Given the multidisciplinary nature of our project, sources from diverse fields, including computer science, marketing, psychology, and data science, are welcomed. Cross-disciplinary insights can enrich our understanding of customer personalization.

Selection Criteria for Exclusion:

1. **Irrelevance to Research Objectives:**

- Sources that do not directly address AI-powered customer personalization, dynamic product recommendations, behavioral segmentation, or real-time customer insights are excluded from consideration.

2. **Outdated Information:**
   - Sources published more than 4 years ago are excluded unless they offer foundational or historical context that is essential for understanding the evolution of the field.
3. **Low Credibility and Unverified Sources:**
   - Non-peer-reviewed sources, self-published materials, and content from unverified or questionable sources are excluded to maintain the integrity of our review.
4. **Unsubstantiated Claims:**
   - Sources lacking empirical evidence, rigorous methodology, or clear support for their claims and findings are excluded.
5. **Off-Topic or Generalized Content:**
   - Sources that provide general information about AI or personalization without specific relevance to our project's focus areas are excluded.
6. **Language and Accessibility:**
   - Sources in languages other than English or those with limited accessibility are excluded unless their inclusion is essential for the completeness of our review.

Data Extraction

In this section, we present a summary of the relevant information extracted from the selected sources. Our data extraction process aimed to capture key findings, methodologies employed, theoretical frameworks discussed, and other essential details to inform our research on the "AI-Powered Customer Personalization Platform."

Source 1: Personalization the artificial intelligence way

- **Publication Details:** Author: Andrew Pearson,
- **Publication date:** August 2020
- **Research Objectives:** In light of these themes, the research objectives for further investigation include assessing AI project success rates, exploring the five types of AI and their applications, analyzing AI's role in diverse industries, understanding customer demand for personalization, and evaluating AI's significance in modern marketing. These objectives will guide our research efforts and inform the development of our "AI-Powered Customer Personalization Platform" project.
- **Methodology & keywords:** personalisation, real-time marketing, artificial intelligence, machine learning, deep learning, lookalike marketing, chatbots, customer lifecycle, emotional recognition, psychometrics, image search, website morphing, voice-assisted search, voice recognition, programmatic advertising.
- Article:

  05_ED_Pearson-12-12-19[1].pdf

Source 2: Recommender System Based on Purchasing History

- **Publication Details:** Authors: Shengshi Yuan, Tiantian Zhang, Xinyi Tan
- **Publication date**: December, 2020
- **Research Objectives:** The Recommendation system has brought huge impacts to the e-commerce industry, and it even influences the business model of those online retail service providers. In this project, they constructed the recommender systems mainly using techniques that belong to collaborative filtering and matrix factorization. The neural network using the user-product pair

performed the best among the algorithms for both user-brand recommendation and user-product recommendation.

- **Methodology & keywords:** Model 1: Item-Based Collaborative Filtering, Model 2: Matrix Factorization with Implicit Alternative Least Squares , Model 3: Neural Network
- Article:

> AI based customer recommandation.docx

Proposed Approaches

1) Rank-Based Product Recommendation

Objective:

- Identify the goods with the best customer reviews and suggest them.
- Market to new consumers with well-liked, rated products.
  Reduce the "Cold Start Problem."

Outputs:

- Provide the top 5 items that have received at least 50/100 ratings or user interactions.

Approach:

1. Determine each product's average rating.
2. Compute the overall rating count for each product.
3. Create a DataFrame and arrange the columns by average ratings.
4. Create a function to return the top 'n' products that satisfy the required minimum level of engagement.

2) Similarity-Based Collaborative Filtering

Objective:

- Offer people individualised recommendations based on their tastes and behaviour.

Outputs:

- Suggest the top 5 products based on user interactions with others who share similar interests.

Approach:

- For simplicity, convert user_id values to integers.
- Use the techniques listed below to find users that share common characteristics:
    - Use cosine similarity to calculate the similarity score between the selected user and all other users.
    - Select the users with the highest similarity scores and isolate them.
    - Remove the initial user's score for similarity from the list.
    - Show the user list with comparable users.
- Make product recommendations by: - Determining the merchandise that the initial user has interacted with (observed_interactions).
    - List 'n' products for each similar user that they have dealt with but the original user hasn't.
    - Offer the predetermined amount of product suggestions.

3) Model-Based Collaborative Filtering

Objective:

- Provide customers with tailored recommendations that take into account their interests and behaviour while solving issues with data sparsity and scalability.

Outputs:

- Provide a user-specific list of the top 5 products.

Approach:

1. To maximise memory use and computational effectiveness, convert the matrix of product ratings into a CSR (Compressed Sparse Row) matrix.
2. Singular value decomposition (SVD) can be used to reduce the number of dimensions to 50 latent features.
3. Multiply the U, sigma, and Vt matrices to determine expected ratings.
4. Make product suggestions based on anticipated ratings:
    - Obtain the user's expected and actual ratings.
    - Create a DataFrame with the product names, actual and projected ratings, and both.
    - Filter the DataFrame to include products without ratings.
    - DataFrame should be sorted by expected ratings.
    - Show the best 'num_recommendations'

Gaps

- **Effective Integration of AI Types:** While the literature discusses various types of AI, such as sound, time series, text, image, and video, there's limited exploration of how these AI types can be effectively integrated into a unified personalization platform. Further research is needed to understand the synergies and challenges of combining these AI modalities to provide a seamless and holistic customer experience.
- **Real-time Personalization Challenges:** Current research emphasizes the potential of real-time customer insights. However, there's a gap in understanding the practical challenges of implementing real-time personalization at scale, including data processing, infrastructure requirements, and ensuring data privacy compliance. Investigating solutions to these challenges is essential for successful real-time personalization.
- **Customer Privacy and Ethics:** The literature acknowledges the importance of customer data privacy and ethical considerations, but there's a need for more in-depth analysis and guidelines on how to navigate these issues in the context of AI-powered customer personalization. Research should focus on balancing personalization benefits with privacy protection to build customer trust.

- **Small and Medium-sized Enterprises (SMEs):** Most studies concentrate on AI applications in large enterprises, leaving a gap in understanding how SMEs can leverage AI for customer personalization. SMEs have unique resource constraints and customer dynamics, requiring tailored solutions and strategies.
- **Long-term Impact of Personalization:** While personalization can enhance short-term customer engagement, there's limited research on its long-term impact on customer relationships and brand loyalty. Investigating how sustained personalization efforts affect customer retention and lifetime value is crucial for businesses.
- **Benchmarking and Evaluation Metrics:** Existing literature lacks standardized benchmarks and evaluation metrics for AI-powered customer personalization. Developing robust metrics and benchmark datasets will facilitate more accurate comparisons of different personalization approaches and their effectiveness.

Conclusion

We presented a systematic comparison of various collaborative filtering systems on a real-life purchase dataset. We showed that recommender systems can be developed on purchase rather than rating data, and that, in such cases, algorithms comparison is different from the standard rating case. In particular, on our data, the simplest algorithm based on bigram association rules obtained the best performances. We believe that such case-studies are necessary to better understand the specificities of purchase datasets and the factors that impact recommender systems for retailers. Our results show that factors such as purchase recency and context-awareness may be at least as important as the choice or the design of a well-performing algorithm. They also show that the relative performances of the algorithms vary depending on the setting to which they are applied, so that all algorithms have to be tested at all stages of the development of the recommender system.

Citation and Referencing

[1] Daqing Chen, Sai Laing Sain, and Kun Guo, “Data Mining for the Online Retail Industry: A Case Study of RFM Model-Based Customer Segmentation Using Data Mining,” (August 27, 2012) Database Marketing & Customer Strategy Management 19, 197-208

[12] Yifan Hu, Yehuda Koren, and Chris Volinsky, “Collaborative Filtering for Implicit Feedback Datasets,” 2008 Eighth IEEE International Conference on Data Mining, 2008. https://doi.org/10.1109/icdm.2008.22.

[3] Yehuda Koren, “Matrix Factorization Techniques for Recommender Systems,” Published by the IEEE Computer Society, IEEE 0018-9162/09, pp. 42- 49, ©IEEE, August 2009.

[4] Michael Kechinov, “eCommerce Events History in Cosmetics Shop,” (March 2020) distributed by REES46 Marketing Platform, https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop