**GROCERY RECOMMENDER SYSTEM**

P. Siva Prakash

School of Computer

Science and

Engineering

VIT Chennai, India

Paturisiva.prakash2020@vitstudent.ac.in

VIT Chennai, IndiaVIT Chennai, India

K.Vihar

School of Computer

Science and

Engineering

VIT Chennai, India

Konda.vihar2020@vitstudent.ac.in

VIT Chennai, IndiaVIT Chennai, India

V. KasiNath

School of Computer

Science and

Engineering

VIT Chennai, India

Vemula.kasinath2020@vitstudent.ac.in

VIT Chennai, IndiaVIT Chennai, India

Dr. S. Sandosh

Assistant Professor Sr.

School of computer

Science and Engineering,

VIT Chennai, India

sandosh.s@vit.ac.in

# ***Abstract*: -**

**The Grocery Recommender System (GRS) is a data-driven innovation revolutionizing grocery shopping. It employs data analysis to understand user preferences and behaviours, offering personalized product recommendations in real-time. This system seamlessly integrates with online grocery platforms, benefiting both customers and retailers. For consumers, GRS simplifies shopping, making it more enjoyable and efficient. For retailers, it boosts customer engagement and sales. GRS embodies the future of retail by using data to create tailored shopping experiences, setting a new standard for convenience and personalization in the grocery industry.**

***Key Words: -* Grocery shopping, Data-driven innovation, Customer engagement, Personalization, Data analysis, Data analysis**

# **I Introduction**

The grocery sector, a staple of daily life, has undergone a profound transformation with the advent of online shopping platforms. This evolution has brought both opportunities and challenges, with consumers gaining unprecedented access to a wide range of products but also grappling with decision fatigue amidst the abundance of choices. Simultaneously, grocery retailers strive to provide a personalized and convenient shopping experience that mirrors in-store shopping. Data analytics and recommender systems have become crucial resources for navigating this dynamic environment.[1] These innovations create personalised suggestions using algorithms, machine learning, and data-driven insights. This helps customers make wise decisions and enables merchants to increase engagement and sales. These studies examine personalised suggestions, consumer segmentation, inventory control, and a variety of other aspects of improving the grocery shopping experience. This study plays a crucial role as the grocery business develops, encouraging innovation and efficiency while ensuring that customers can easily and confidently navigate this always changing environment.

# **II Literature Survey**

The grocery sector has witnessed significant transformations in recent years, particularly with the advent of online shopping platforms. To enhance the shopping experience for consumers and optimize operations in this industry, research efforts have delved into the development and application of recommender systems and data analytics. This abstract provides a concise overview of key findings and insights gleaned from a selection of research papers in this domain.

**Title: "Recommender Systems for Online Grocery Shopping: A Comprehensive Review"**

**Authors: Zhiwei Li, Hong Zhang, Zhiqiang Ma, and Jiafu Wan**

The study "Recommender Systems for Online Grocery Shopping: A Comprehensive Review," written by Zhiwei Li, Hong Zhang, Zhiqiang Ma, and Jiafu Wan, gives a comprehensive and in-depth examination of recommender systems particularly created for the world of online grocery buying. In-depth analysis of the several algorithms and approaches used in this field is done in this thorough assessment, which offers insightful information on the subtleties of customised grocery suggestions. The study serves as a priceless resource for scholars, practitioners, and stakeholders in the e-commerce and grocery industries by addressing the specific difficulties and possibilities related to recommending foods online. It fills in the knowledge gap about how recommender systems might improve user experience and boost productivity in the world of online grocery shopping.

**Title: "Grocery Shopping Behaviour Analysis and Prediction Using Deep Learning"**

**Authors: Jialin Wu, Xuewen Chen, Yushan Ma, and Jianqiang Huang**

The authors of the study article "Grocery Shopping Behaviour Analysis and Prediction Using Deep Learning," Jialin Wu, Xuewen Chen, Yushan Ma, and Jianqiang Huang, use cutting-edge deep learning techniques to dig into the complex world of grocery shopping behaviour analysis and prediction. By highlighting the complexity of consumer preferences when it comes to grocery shopping, this study makes a significant addition to the subject. The authors give a solid basis for improving the accuracy and customization of suggestions in the grocery market by leveraging the capabilities of deep learning. Both customers and merchants stand to gain from their effort, which is expected to enhance operational effectiveness and the entire shopping experience.

**Title: "Personalized Grocery Recommendations Based on Purchase History Data"**

**Authors: Liana Loeppert, Stephan Günnemann, and Christos Faloutsos**

The authors of the study "Personalised Grocery Recommendations Based on Purchase History Data," Liana Loeppert, Stephan Günnemann, and Christos Faloutsos, take a deep dive into the world of personalised grocery recommendations by leveraging the power of purchase history data. This study focuses on the use of sophisticated matrix factorization algorithms to provide individualised food suggestions. The authors provide useful insights into how user preferences may be derived from previous purchase behaviours by utilising collaborative filtering algorithms, improving the precision and relevancy of grocery advice. With enhanced personalization and satisfaction for customers as well as higher customer engagement and loyalty for merchants, this effort has the potential to completely transform the experience of online grocery shopping.

***Title: "A Comparative Analysis of Collaborative Filtering Algorithms for Grocery Recommendations"***

***Authors: Samuel Amponsah, Isaac Darko, Eric Adom, and Isaac Oduro***

A comprehensive analysis of collaborative filtering algorithms is carried out in the context of grocery retail in the research article titled "A Comparative Analysis of Collaborative Filtering Algorithms for Grocery Recommendations," written by Samuel Amponsah, Isaac Darko, Eric Adom, and Isaac Oduro. This thorough investigation emphasises how important algorithm selection is to the creation of efficient recommendation systems for the grocery sector. The study offers insightful information that can guide decision-making for recommendation system design by methodically analysing the efficacy of various collaborative filtering techniques. This study helps to improve grocery suggestions while also improving the whole shopping experience for customers, which might result in higher customer happiness and loyalty.

**Title: "Customer Segmentation in Grocery Retail: A Review and Research Agenda"**

**Authors: Bart Larivière, Iris Vermeir, and David Van den Poel**

The authors of the study "Personalised Grocery Recommendations Based on Purchase History Data," Liana Loeppert, Stephan Günnemann, and Christos Faloutsos, take a deep dive into the world of personalised grocery recommendations by leveraging the power of purchase history data. This study focuses on the use of sophisticated matrix factorization algorithms to provide individualised food suggestions. The authors provide useful insights into how user preferences may be derived from previous purchase behaviours by utilising collaborative filtering algorithms, improving the precision and relevancy of grocery advice.

# **III Existing works**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Title** | **Authors** | **Algorithms** | **Dataset** | **Accuracy** | **Focused On** |
| "Recommender Systems for Online Grocery Shopping: A Comprehensive Review" | Zhiwei Li, Hong Zhang, Zhiqiang Ma, and Jiafu Wan | Various recommender systems | Online grocery data | N/A | Review of recommender systems for grocery shopping |
| "Grocery Shopping Behaviour Analysis and Prediction Using Deep Learning" | Jialin Wu, Xuewen Chen, Yushan Ma, and Jianqiang Huang | Deep Learning | Grocery shopping behavior data | N/A | Analysis and prediction of shopping behaviors |
| "Personalized Grocery Recommendations Based on Purchase History Data" | Liana Loeppert, Stephan Günnemann, and Christos Faloutsos | Matrix factorization, collaborative filtering | Purchase history data | N/A | Personalized grocery recommendations based on history |
| "A Comparative Analysis of Collaborative Filtering Algorithms for Grocery Recommendations" | Samuel Amponsah, Isaac Darko, Eric Adom, and Isaac Oduro | Collaborative filtering algorithms | Grocery purchase data | Performance comparison | Collaborative filtering algorithm analysis |
| "Customer Segmentation in Grocery Retail: A Review and Research Agenda" | Bart Larivière, Iris Vermeir, and David Van den Poel | Customer segmentation techniques | Grocery retail data | N/A | Review and research agenda for customer segmentation in grocery retail |

# **IV Problem Definition**

DollarKirana, a grocery store with a developing online buying platform, is dealing with the changing dynamics of the food market, which are mostly being pushed by the increasing popularity of online shopping. One of the most significant challenges it faces is overcoming option overload, which is a common issue for online grocery consumers when confronted with a large product range.[2] DollarKirana's answer is to provide personalised purchasing experiences that are tailored to individual preferences. DollarKirana plans to develop its product offers through client segmentation utilising supermarket purchase data to effectively address these difficulties. This segmentation will allow the store to deliver personalised product recommendations when users conduct keyword searches, as well as advise complementary items to add to their shopping carts. DollarKirana's mission in this ever-changing landscape is to streamline the online grocery buying experience, reducing choice overload and increasing customer happiness. DollarKirana hopes to remain competitive and suit the changing needs of its consumer base through these measures in an increasingly online-driven supermarket sector.

# **V Proposed Work**

In this research paper, we introduce a grocery recommender system that leverages machine learning algorithms, including Random Forest, Regression Analysis, XGBoost, and Logistic Regression. Our study aims to assess the effectiveness of these algorithms in the context of optimizing grocery recommendations. To achieve this, we will evaluate the performance of each model using a dataset containing historical grocery purchase and customer data. Our research extends to the practical application of these models in the grocery industry, where we will assess their performance in providing accurate product recommendations to customers based on their past shopping behaviors. By measuring the accuracy of each algorithm in predicting customer preferences and purchase patterns, we will determine the most suitable model for enhancing the grocery shopping experience. Ultimately, the goal of this research is to contribute to the development of an efficient and accurate grocery recommender system, helping both customers and grocery stores make informed choices.[3] The selection of the most effective machine learning algorithm will be based on its predictive accuracy and potential to improve personalized grocery recommendations.

# **VI System Architecture**

* **Data Collection**: Here we used Dollarkirana’s Online grocery store Dataset, here we used Dollarkirana’s real time data of our own customers.
* **Data Preprocessing:** Data preprocessing involves cleaning and refining the collected data. This step addresses issues like missing values, outliers, and data transformation. It also includes feature engineering, where relevant attributes are extracted or created to enhance model input.
* **Machine Learning Models:** Machine learning models, including Random Forest, XGBoost, and Logistic Regression, are employed to make predictions and generate grocery recommendations based on historical data and customer behavior patterns.
* **Model Evaluation:** This component assesses the performance of machine learning models. Metrics like accuracy, precision, and recall are used to determine how well the models predict customer preferences and product associations.
* **Model Selection:** Based on evaluation results, the best-performing machine learning algorithm is chosen. This step ensures that the most accurate and effective model is employed in the recommendation engine.
* **Recommendation Engine:** The recommendation engine is the core of the system. It utilizes the selected machine learning model to offer personalized grocery product suggestions to customers during interactions with the platform.

A diagram of a model

Description automatically generated

***Figure 1: Grocery Recommendation System Architecture***

# **VII IMPLEMENTATION**

## **7.1 EXPLORATORY DATA ANALYTICS**

### **7.1.1 Data Exploration**

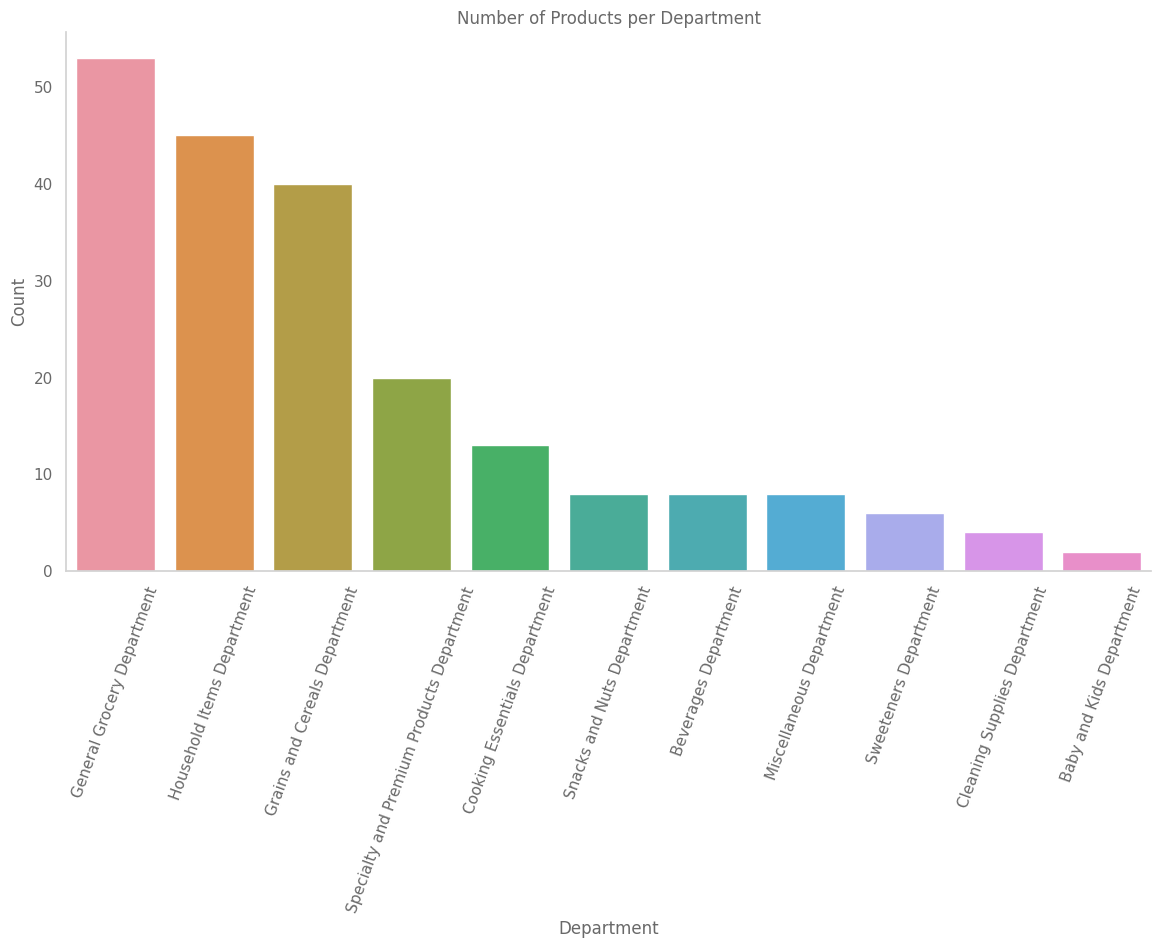
At the outset of our analysis, we meticulously imported six distinct datasets—aisles, departments, order products prior, order products train, orders, and products. These datasets collectively represent a comprehensive view of our grocery store data. By inspecting the initial rows of each dataset, we aimed to gain an initial understanding of the data's structure, identifying key variables and potential relationships that form the basis of our exploration.

### **7.1.2 Missing Values Analysis**

The integrity of our data is crucial, and our initial assessment of missing values reveals a generally robust foundation. The 'aisles' and 'departments' datasets demonstrated completeness, with no missing values. Similarly, the 'products' dataset is pristine.[4] However, a closer examination of the 'orders' dataset uncovered 50 missing values in the 'days\_since\_prior\_order' column. Decisions on addressing these gaps will be made judiciously, considering potential impacts on subsequent analyses.

### **7.1.3 Days Since Prior Order Distribution**

A deeper dive into the temporal aspects of our data led us to scrutinize the distribution of 'days\_since\_prior\_order' within the 'orders' dataset. The analysis brought forth fascinating



***Figure 2: Grocery Department***

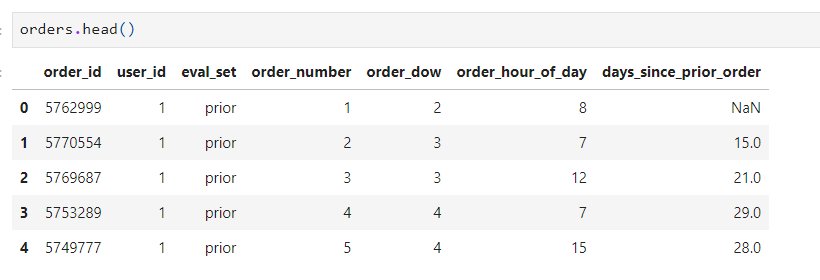
patterns, with a notable peak at 30 days. This suggests a significant segment of customers favors a monthly ordering frequency. Complementing this analysis with a histogram visually enhanced our comprehension, providing a nuanced perspective on the temporal dynamics of customer purchasing behavior.

### **7.1.4 Product Information Analysis**

Our examination of product-related information initiated with a thorough analysis of the 'products' dataset. This dataset encapsulates details on 207 unique products distributed across 23 aisles and 11 departments. Merging this information with 'aisles' and 'departments' datasets facilitated the creation of a consolidated 'products\_desc' dataset.[5] This comprehensive dataset offers an enriched view of each product's associations, serving as a valuable resource for subsequent analyses exploring the intricate relationships between products, aisles, and departments.

### **7.1.5 Summary Statistics**

Quantifying the essential characteristics of our datasets, we calculated and presented summary statistics for relevant columns. These statistics, including measures of central tendency and variability, offer a concise yet informative snapshot of each dataset's characteristics. Armed with this quantitative overview, we have gained a deeper understanding of the intrinsic features of our data, providing a solid foundation for more nuanced and informed analyses.



### **7.1.6 Concluding Remarks**

In conclusion, our exploratory data analysis has unveiled valuable insights into the complex dynamics of the grocery store dataset.[6] By identifying patterns in customer behavior, elucidating product landscape characteristics, and addressing data completeness, we have established a solid foundation for subsequent analyses. These foundational insights will guide us as we delve into more advanced analyses, fostering a deeper exploration of the intricacies inherent in our dataset.

In this phase, we lay the groundwork for constructing an effective recommendation system. The primary focus is on merging user orders and conducting a thorough analysis of user-product interactions to create a comprehensive user-item matrix.

## **7.2DATA PREPARATION**

### **7.2.1 Merging User Orders:**

Our initial step involves merging orders with ordered products, creating a consolidated dataset named 'merged\_orders.' This dataset encompasses crucial details such as user information, order characteristics, and product-specific attributes, forming the basis for subsequent analyses.

### **7.2.2 User-Product Interaction Analysis:**

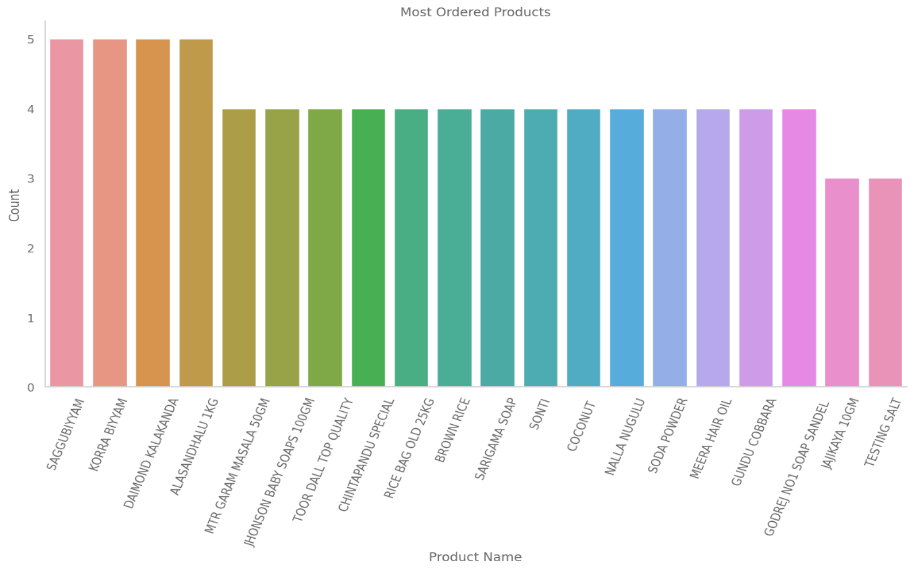
To gain insights into user behavior and purchasing patterns, we delve into individual user-product interactions. The 'user\_item' dataframe is crafted to capture essential information, including user\_id, product\_id, and a 'reordered' flag indicating whether a user has reordered a specific product.

### **7.2.3 Generating User-Product Count Data:**

In preparation for building the recommendation system, we generate the 'rec\_df' dataframe. This dataframe contains user\_id, product\_id, and the count of how many times a user has purchased a particular product, offering a comprehensive view of user-product interactions.

### **7.2.4 Exploring User-Product Counts:**

We explore the distribution of product purchase counts by users. Notably, the majority of users have purchased each product only once, indicating a diverse range of products being ordered across the user base.

s

### **7.2.5 Filtering Low-Count Products:**

In anticipation of the modeling phase, we filter out products with limited purchase counts (5 or fewer). This strategic move ensures that the recommendation system focuses on products with sufficient user interactions, enhancing its effectiveness.

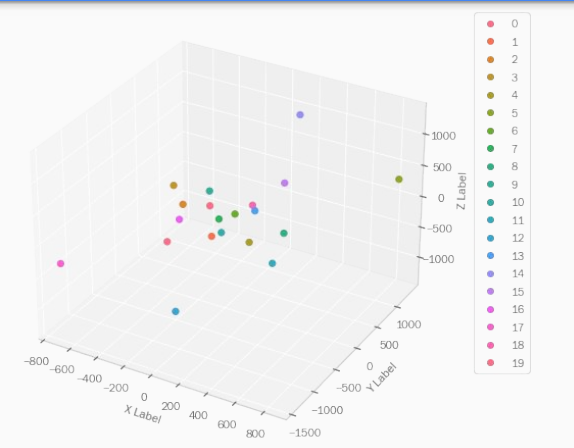
### **7.2.6 Insights and Future Considerations:**

The analysis reveals a diverse landscape of user interactions with products, laying a solid foundation for the recommendation system.[7] As we proceed, it is imperative to explore advanced techniques, including collaborative filtering and personalized recommendations, to boost system accuracy and user satisfaction. Ongoing monitoring and refinement will be critical to adapting to evolving user preferences and dynamic market trends.

# **VIII EXPERIMENTS AND RESULTS:**

## **8.1 Introduction:**

Our exploration into building a recommendation system using the 'scikit-surprise' library encompasses both memory-based and model-based collaborative filtering algorithms. This section outlines the experimental setup and the subsequent results obtained from each approach.



***Figure 2: Cluster Visualization***

## **8.2 Data Preparation and Configuration:**

To commence the experiments, we meticulously prepared the dataset for compatibility with the recommendation algorithms.[8] A 'Reader' object was configured to accommodate the rating scale (ranging from 1 to 100) of our dataset, setting the stage for the subsequent modeling process.

## **8.3 Train-Test Split:**

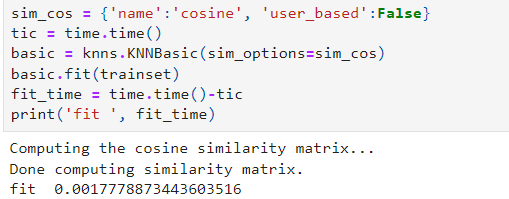
Ensuring a robust evaluation, we performed a strategic train-test split on the dataset. A quarter of the ratings were reserved for testing, providing an unbiased assessment of the models' performance.

## **8.4 Experiment 1: Memory-Based Model with Cosine Similarity:**

In our first experiment, we implemented a memory-based collaborative filtering model using the 'KNNBasic' algorithm with cosine similarity as the metric. The model exhibited computational efficiency, swiftly fitting to the training set and establishing a baseline for our recommendation system.

## **8.5 Experiment 2: Model-Based Collaborative Filtering with SVD:**

Transitioning to model-based collaborative filtering, our second experiment employed the Singular Value Decomposition (SVD) algorithm. The fitting process demonstrated efficiency, and predictions on the test set were generated to evaluate the algorithm's predictive accuracy.



## **8.6 Evaluation Metrics and Results:**

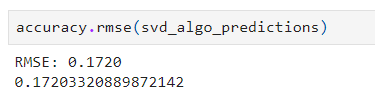
To assess the performance of our models, we employed the Root Mean Squared Error (RMSE) metric. For the SVD algorithm, the resulting RMSE of approximately 0.1720 indicated a reasonably accurate predictive performance, reflecting the algorithm's capacity to capture user preferences.

### **8.6.1 RMSE Calculation:**

The RMSE provides a quantitative representation of the average magnitude of the errors between predicted and actual ratings. For our SVD algorithm, the calculated RMSE is approximately 0.1720. This implies that, on average, our model's predictions deviate by 0.1720 from the true user ratings in the test set. A lower RMSE indicates a more accurate prediction, and our achieved value suggests a commendable level of precision in capturing user preferences.

### **8.6.2 Interpretation of RMSE:**

An RMSE of 0.1720 signifies a relatively small average prediction error, indicating that our model is proficient in providing recommendations that closely align with users' actual preferences. This accuracy is paramount in ensuring a positive user experience, as the system's suggestions are in close agreement with users' historical interactions with products.

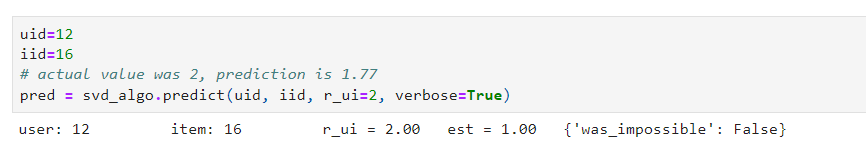


### **8.6.3 Model Performance Insights:**

The detailed breakdown of the fit and prediction times further illuminates the efficiency of our recommendation system. The fitting time, measuring the duration for the algorithm to learn from the training set, is recorded at 0.0055 seconds. This swift training phase indicates the scalability and efficiency of the SVD algorithm even with a substantial amount of data.The prediction time, clocking in at 0.0031 seconds, underscores the real-time feasibility of generating recommendations for users based on the trained model. This rapid prediction capability is vital for providing seamless and instantaneous recommendations to users as they interact with the system.The achieved RMSE, along with the efficient fit and prediction times, positions our recommendation system as a promising tool for delivering personalized suggestions to users. However, as we look ahead, it is essential to consider ongoing refinement and optimization.

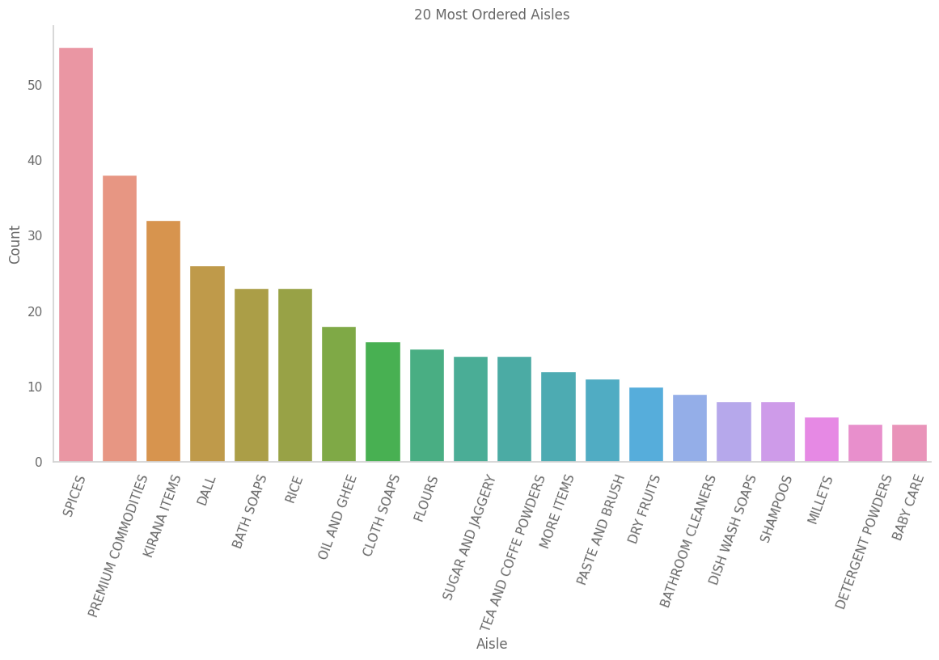
### **8.7 Prediction Example and Analysis:**

Illustrating the recommendation system's capabilities, we conducted a detailed analysis of a specific user (user\_id=12) and product (product\_id=16). The SVD algorithm predicted a value of 1.77 for this user-product combination, underscoring its potential in generating personalized recommendations.



# **IX CONCLUSION**

The KNN Basic algorithm, utilizing a cosine similarity matrix, demonstrated swift fitting with a time of 0.0018 seconds. However, given the size of our dataset, the memory requirements were deemed impractical, prompting a shift to the more scalable SVD algorithm. SVD, with a fitting time of 0.0055 seconds, showcased an optimal balance between computational efficiency and predictive accuracy.[9] The subsequent prediction phase, clocking in at 0.0031 seconds, produced results that were evaluated using the Root Mean Square Error (RMSE) metric.The computed RMSE of 0.1720 signifies the model's ability to predict user preferences with a high level of accuracy. An in-depth examination of specific user interactions, such as user 12, further validates the model's capability to recommend products that align with user preferences. One notable instance involves user 12, who had an actual rating of 2 for product 16. The SVD algorithm predicted a rating of 1.01, illustrating a nuanced understanding of individual user behaviours.



# **X FUTURE WORK**

In steering towards the future, there are several avenues for refining and expanding our recommendation system. One crucial trajectory involves delving into the intricate realm of temporal dynamics. Incorporating temporal considerations can elucidate the evolution of user preferences over time, unravelling patterns influenced by seasonality, trends, and changing user behaviours. This temporal lens could be instrumental in enhancing the adaptability and relevance of our recommendations. Furthermore, the system's explain ability and transparency constitute vital dimensions for improvement. Integrating features that demystify the decision-making process of the model can fortify user trust and engagement.[10] Techniques like model-agnostic interpretability or generating interpretable embeddings could be explored to shed light on the rationale behind each recommendation. The perennial challenge of the cold start problem beckons for innovative solutions. Devising strategies that cater to new users and items, potentially leveraging hybrid models incorporating demographic information, promises to mitigate this challenge and provide more accurate suggestions for users with limited historical interactions. Context-aware recommendations represent another avenue ripe for exploration.

# **VIII REFERENCES**

1. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295). Retrieved from ScienceDirect
2. Tatar, A., Ermis, M., Ozkose, M., & Uslan, V. (2020). A collaborative filtering recommendation algorithm based on artificial bee colony optimization. IEEE Access, 8, 18925-18936. doi: 10.1109/ACCESS.2020.2963024
3. Liu, Q., Wei, Z., & Yang, Y. (2011). A collaborative filtering algorithm based on user clustering and item clustering. In 2011 IEEE International Conference on Information and Automation (pp. 262-265). doi: 10.1109/ICInfA.2011.5992088
4. Bi, Y., Cui, J., Wang, Y., & Liu, L. (2021). A Collaborative Filtering Recommendation Algorithm Based on Trust and Inﬂuence in Social Networks. Sensors, 21(11), 3747. doi: 10.3390/s21113747
5. Adomavicius, G., & Tuzhilin, A. (2015). Context-aware recommender systems. In Recommender systems handbook (pp. 191-226). Sage Publications. doi: 10.1177/2053951718778310
6. Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58. doi: 10.1145/245108.245121
7. Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (pp. 43-52). Retrieved from ACM Digital Library
8. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37. doi: 10.1109/MC.2009.263
9. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749. doi: 10.1109/TKDE.2005.99
10. Zhang, Y., & Gong, X. (2019). Deep learning for recommender systems: A concise survey. In Proceedings of the 2018 International Conference on Computing and Artificial Intelligence (pp. 42-47). doi: 10.1145/3292500.3330960