**Prepared For :**
DataThon 2023

**HO CHI MINH**
November, 19th 2023

# PRODUCT
# DATASET 2

# OUR PROPOSAL

## SALES AND INVENTORY DATA OF VIETNAM RETAILERS

## INTRODUCTION

Data on sales and inventory at retailers in Vietnam is becoming increasingly important and complex. Effective management of this data can help businesses optimize sales processes, predict market trends, and mitigate risks associated with inefficient inventory.

Specifically, data on sales and inventory can be used to:

- Make recommendations for which products are likely to sell in the present and future, and which products have the potential to generate optimal profits for the business.
- Develop products based on feedback from consumers and available data.
- Build programs to predict and recommend products to appropriate users.
- Develop solutions for how to deal with inventory items that are not selling.
- Automate the integration of data into a database to facilitate the management of available data.
- Develop a website platform that facilitates interaction with the system.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## PROBLEM STATEMENT

Proposed solutions:

- Segment customers clearly to address the problem of sold and unsold products.
- Provide appropriate recommendations to customers with data prioritized on sales and unsold products.

Challenges:

- Lack of interaction and user feedback.
- Facing changes in trends and customer requirements.
- Challenges in dealing with new products.
- Excessive subjectivity about the data.
- Inefficient inventory management.
- Difficult to predict shopping due to shortcomings in product forecasting.
- Promotional and discount strategies are not effective.

# OUR PROPOSAL

## SOLUTION OVERVIEW

1.Algorithm Applied:

Hybrid Recommenders: Hybrid Filtering is the combination of two algorithms, Content-based Filtering and Collaborative Filtering. Hybrid Filtering is applied flexibly when the Collaborative Filtering system lacks user behaviors (ratings). In such cases, the system utilizes Content-based Filtering, and conversely, when Content-based Filtering lacks essential features for evaluation, the system resorts to Collaborative Filtering as a substitute.

- Content-based Filtering:

In the content-based method, the system evaluates the characteristics of recommended items. It suggests items based on user profiles or the content attributes of items similar to those chosen by the user in the past. For example, if a person enjoys consuming oranges, the system recommends a similar fruit, such as grapefruit. This approach requires organizing items into groups or identifying the features of each item.

# OUR PROPOSAL

## SOLUTION OVERVIEW

- Collaborative Filtering:
This algorithm predicts a user's preference for an item based on other users who are "similar" to the current user. Determining the "similarity" between users can be based on their level of interest (rating) in other items that the system has known in the past.

2. Why Use Content-based Filtering and Collaborative Filtering?
Because our available data is suitable for the input requirements of both Content-based Filtering and Collaborative Filtering algorithms:

*Content-based Filtering
a. Create product profiles:
- Utilize detailed information from the "MasterData" dataset to create profiles for each product, including attributes such as color, size, category, selling price, purchase price, style, options, brand name.
- Represent these attributes as numerical vectors to calculate similarity between products.

## SOLUTION OVERVIEW

b. Calculate Similarity:

Utilize methods such as cosine similarity to measure the similarity between products based on their profiles.

Suggest inventory items that share attributes with popular or historically user-relevant products.

*Collaborative Filtering:

a. Determine user similarity:

Use information from the "Inventory_ snapshot_ data" dataset to ascertain the level of interest (rating) that users have shown for purchased products.

Base user similarity on this information.

b. Predict user preferences:

Based on user similarity, predict the preferences of a user for unpurchased products.

Recommend inventory items based on predicted user preferences.

# OUR PROPOSAL

## SOLUTION OVERVIEW

3. Output Results:
The system's output will present a list of recommended products for the user based on inventory information and product attributes.

4. Explanation of Algorithm Results:
Combine the recommendation lists from both Content-based and Collaborative Filtering to provide a prioritized and diverse final list. Different weights can be applied to each method based on specific requirements.

Benefits:
Combine the advantages of both methods to offer accurate and diverse recommendations.
Adjust the prioritization between product information and user information.

Evaluate system performance using metrics such as precision, recall, F1-score, or real user evaluation methods.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## METHODOLOGIES

How the Algorithm Operates:

Content-based Filtering:

1. Create Product Profiles: For each product i, create a vector Pi containing attributes such as color, size, selling price, etc.
2. Calculate Similarity: Use a similarity measurement method, such as cosine similarity, to measure the similarity between the profile of the evaluated product Pi and the profile of the user U:

$$\mathrm{milarity}(U, P_i) = \frac{U \cdot P_i}{\|U\| \cdot \|P_i\|}$$

In which vector multiplication is employed, || || denotes the Euclidean length of the vector.

3. Product Recommendations:

$$\mathrm{Prediction}(U, P_j) = \sum_i \mathrm{Similarity}(U, P_i) \times \mathrm{Rating}(i, P_j)$$

Where Rating (i, U) is the rating that user U has given to product i.

# OUR PROPOSAL

## METHODOLOGIES

Collaborative Filtering:

1. Determine Similarity Between Users:

Utilize information from dataset B to determine the level of interest (rating) that users have shown for purchased products. Employ a similarity measurement method, such as cosine similarity, between two users Ui and Uj.

$$\text{Similarity}(U_i, U_j) = \frac{U_i \cdot U_j}{\|U_i\| \cdot \|U_j\|}$$

2. Predicting Preferences:

Based on the similarity between users, predict the preferences of a user for an unpurchased product Pj.

$$\text{Prediction}(U_i, P_j) = \frac{\sum_{U_j} \text{Similarity}(U_i,U_j) \times \text{Rating}(U_j,P_j)}{\sum_{U_j} \text{Similarity}(U_i,U_j)}$$

In which, Rating(Uj, Pj) is the rating that user Uj has given to the product Pj.

Using Machine Learning (ML) is an effective method to leverage both Content-based Filtering and Collaborative Filtering algorithms.

# OUR PROPOSAL

## CORE FUNCTIONALITY

Identifying products that can be sold in the upcoming month as well as those that should be introduced for targeting purposes in prediction is crucial for providing businesses with insights into product viability.

The returned results consist of a list of customers and corresponding products. These products are the ones the business needs to recommend to each customer. With this analysis, the business can optimize its economic activities.

Providing a flexible, cost-effective, and easily deployable solution to help retailers enhance business management.

Suggesting products for businesses based on both low and high interest, allowing for a balanced consideration when sourcing these items.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## PERFORMANE METRICES

Content-based Filtering: Precision and Recall

Precision: The ratio of the number of recommended products that users are genuinely interested in to the total number of recommended products.

Recall: The ratio of the number of recommended products that users are genuinely interested in to the total number of products that users are genuinely interested in.

+F1 Score: Combines both precision and recall into a single number, the harmonic mean of the two.

Mean Squared Error (MSE): Measures the magnitude of the difference between predicted and actual ratings.

Collaborative Filtering:
+Root Mean Squared Error (RMSE): Measures the average difference between predicted and actual ratings. Lower values indicate higher model performance.

+Mean Absolute Error (MAE): Measures the average magnitude of the difference between predicted and actual ratings.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## PERFORMANE METRICES

Content-based Filtering: Precision and Recall
+Precision and Recall (for product recommendation systems):
Similar to Content-based Filtering, help evaluate the accuracy and coverage of the system.
+Similarity Metrics:

Measure the similarity between users or products, providing information about the quality of the model.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## TIMELINE AND ROADMAP

1h => 2h: Project kickoff and Strategy

3h => 5h: Deep understanding data, ELT data

6h => 11h: Data visualization, database construction, data warehousing, and providing advanced predictions about desired products.

12h => 21h: Optimize processing based on the model to make predictions using problem-solving algorithms for a user-friendly recommendation system (ML, DL).

22h => 24h: Operational testing and addition of necessary features to evolve into a practical product for businesses (potentially developed into a web or app based on data).

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## USERINTERFACE OR INTERACTION

Businesses can rely on the dashboard to visualize which products will be the focal point and which products are likely to be sold. The system helps identify products with inventory risks to meet business demands.

From the data, we conduct analysis. When a new product emerges, the system recognizes it based on a set of attributes and proposes whether the business should consider importing this new source of goods or not.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## LIMITATIONS AND FUTURE ENHANCEMENTS

-Challenges in Determining Interest:
If there is no rating information, determining the level of user interest in a product becomes challenging.

Collaborative Filtering models often rely on ratings to measure the similarity between users or products.

-Capability to Synthesize Limited Interactions:
If the actual number of interactions is significantly less than the total number of users and products, the model may struggle to accurately identify similarities.

-Cold Start Risk:
If a new user or new product appears without any ratings, Collaborative Filtering models may face challenges in making effective predictions.

-Interaction Assumption:
Collaborative Filtering assumes that there is interaction between users and products to measure similarity. Without ratings, this assumption becomes weak.

# LIMITATIONS AND FUTURE ENHANCEMENTS

1. Additional Rating Solution:
Data Collection Enhancement:

Strengthen data collection by employing various methods to encourage users to rate products, such as posing questions, creating short surveys, or providing incentives for users who submit reviews.

Utilizing Machine Learning Models for Rating Prediction:

Employ machine learning models (e.g., regression models) to predict ratings based on other information, such as product attributes, user shopping history, or other relevant features.

Encouraging Ratings through Gamification:

Implement gamification to motivate users to rate products. For example, create different levels, badges, or rewards based on user product reviews.

# OUR PROPOSAL

## LIMITATIONS AND FUTURE ENHANCEMENTS

2. Business Optimization Solutions
Algorithms

+Addressing Cold Start:
Utilize Content-based methods or gather feedback information from new users to tackle the Cold Start issue for new users or products.

+Using Hybrid Models:
Combine both Collaborative Filtering and Content-based Filtering to leverage the advantages of both methods. This helps minimize the impact of the missing rating issue.

+Parameter Optimization:
Continuously optimize the parameters of the Collaborative Filtering model to ensure optimal performance on real-world data.

+Employing Deep Learning Models:
If there is sufficient data, consider deploying deep learning models for Collaborative Filtering to leverage the ability to learn complex patterns without solely relying on ratings.

+User Care and Interaction:
Enhance user care and interaction to gain a better understanding of preferences and individual tastes, thereby improving the recommendation capabilities of the system.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## CONSOLUTION

After the successful completion of the project, there could be a framework to predict essential products for the business:

+ Meeting all the specified requirements
+ Providing a solution to the business problem
+ An app/website that utilizes data to make predictions for the business department.

The team's completed MVP can be used to predict products for users that are needed and relevant to what they are searching for, along with some related options.

HO CHI MINH
NOVEMBER, 19th, 2023

# OUR PROPOSAL

## REFERENCES


1."https://viblo.asia/p/xay-dung-content-based-filtering-rs-recommender-system-co-ban-phan-2-bWrZnVovZxw"

2. "https://viblo.asia/p/xay-dung-collaborative-filtering-rs-recommender-system-co-ban-phan-3-Az45bMqolxY"

3. "https://viblo.asia/p/xay-dung-content-based-filtering-rs-recommender-system-co-ban-phan-2-bWrZnVovZxw"

4."https://www.researchgate.net/publication/349906752_Research_on_Hybrid_Collaborative_Filtering_Recommendation_Algorithm_Based_on_the_Time_Effect_and_Sentiment_Analysis"


HO CHI MINH
NOVEMBER, 19th, 2023

# CONTACT US

**Address**

hoduytruong280220@gmail.com

**Phone Number:** 0397395029

**Thank you so much!!!**

HO CHI MINH
NOVEMBER, 19th, 2023