
# Full Report

# Recommendation AI-enabled system for local store

By

675380020-2 Mr. Thant Lwin Oo

675380013-9 Mr. Somkiad Khenwiset

Present

Asst. Prof. Dr. Chitsutha Soomlek

SC348 810 Software Development and Project Management for Data Science and Artificial Intelligence

Semester 1/2567

College of Computing

Khon Kaen University

**Declaration: Generative AI is used to translate documents, extract knowledge from documents and implement consultation.**

1. **Description of the Application and Identification of the country where the product will be marketed and sold**

1.1. Application: Recommendations System

The application is an AI-based recommendation system specifically designed for local convenience stores. Its primary function is to analyze historical purchase data to identify patterns and associations between products that customers frequently buy together. Leveraging this data, the system generates real-time recommendations for store employees and managers to optimize shelf arrangements, ensuring related products are placed in proximity for easier to get them together. Additionally, it assists cashiers in suggesting complementary products to customers at the point of sale. This system aims to enhance the customer shopping experience and has the potential to increase overall sales by promoting additional products that customers may not have initially considered.

1.2. Country: Thailand

This AI system is intended for deployment in Thailand, with a focus on local convenience stores across various regions. Each region in Thailand has its own unique culture that influences buying behavior. The system will be particularly beneficial for convenience stores by providing insights into the products that specific populations frequently purchase together. This allows store managers to make informed decisions about inventory management and product placement, adapted to the preferences of their local customer base.

2. **Key Dimensions of Recommendation System based on Thailand AI Policy Framework**

2.1. Privacy

The AI system should be designed to protect personal data, ensuring that no sensitive information is misused. Compliance with Thailand's Personal Data Protection Act (PDPA) is mandatory, which involves obtaining explicit consent from owner for data collection, anonymizing data wherever possible, and securely managing data storage and access.

According to this dimension, the recommendation system will use the data from transactions of POS system, which the data will be lists of products that customers purchase together in those transactions, which means that all the data that system collected will anonymously customer data.

2.2. Security

AI systems should incorporate robust security measures to protect against data breaches, unauthorized access, and cyber threats. This includes encryption of data at rest and in transit, regular security audits, and the implementation of access controls to limit who can view or

modify sensitive data. The system should also have mechanisms to quickly detect and respond to any security incidents.

According to this dimension, the recommendation system will be deployed in offline environments which the person that can access the data is the manager and employees of those stores. Also implement authentication to limit access to model results and logs.

2.3. Reliability

AI systems should establish trust in AI systems by ensuring accurate, reliable, and reproducible results. Maintaining data quality control is crucial to prevent errors that could damage the credibility of AI systems, especially given the uncertainty about the potential impacts of AI decisions.

According to this dimension, the recommendation system will use an algorithm that that have certain metric to evaluate the accuracy of model and can reproducibly result in the same way as long as dataset still same.

2.4. Fairness

AI system should be designed and used to promote fairness, equality, and inclusivity, ensuring that all groups in society, especially the disadvantaged, benefit equally, without bias, discrimination, or creating social inequality.

According to this dimension, the recommendation system will show the result based on what its algorithm has learned, which the result depends on the data in those stores the result still need human to make decision.

2.5. Transparency

AI systems should be designed to ensure that humans can understand and monitor their operations, including how data is used and how decisions are made. This involves making AI processes explainable, traceable, and auditable to build trust and accountability. Additionally, explanations of AI capabilities and limitations should be communicated in a timely and appropriate manner based on the expertise of the audience.

According to this dimension, the recommendation system will use algorithms that can explainable such as Apriori or FP-Growth algorithms which use for finding association rules in machine learning to build the model which have effectiveness metrics to evaluate the model and can be interpreted.

2.6. Accountability

AI systems should include mechanisms to ensure accountability and clear traceability for any impacts caused by AI, with provisions for addressing or taking responsibility for potential damages. Additionally, all stakeholders involved in AI should engage in consultations and plan for managing long-term risks and impacts.

According to this dimension, the recommendation will have log to keep the model result to trace the model degradation, and the stakeholder such as business owner and manager can consult to developer support when has a problem.

2.7. Human-Centricity

AI systems should prioritize a human-centric approach, ensuring human control over critical decisions. AI should be developed to benefit humanity, avoid harm, promote human values, and contribute to sustainable development for society and the environment.

According to this dimension, recommendation system will show the results of model to user for help them make decisions, which mean the final decision would come from human.

**3. Stakeholders**

3.1. Store Owners

They decide to invest in this system. If they are interested in how the system can enhance their business operations, improve customer satisfaction, and ultimately increase sales. And they are crucial to ensure that the system complies with AI policy.

3.2. Store Employees

The primary end users, particularly managers use system recommendation to plan the shelves arrangement in the store, cashiers and sales assistants interact with the system daily and directly use the system result.

3.3. POS System Providers

These providers are responsible for integrating the AI system with the existing Point of Sale (POS) systems in stores.

3.4. Developers

They will respond to developments of this system and ensure that system will seamlessly deploy in the stores system.

3.5. Customer

They do not directly use the system, but they are recipients of system recommendations. If the system works effectively, it will improve their shopping experience.

**4. Metrics use to Monitor and AI Policy Compliance**

4.1. Privacy

**Metric:** Percentage of customer data that is anonymized before being processed.

**Measure:** How complete that data was anonymized by using preprocess log compare with total data to ensure that the system complies with Thailand's Personal Data Protection Act.

$$\text{Anonymized Percentage} = \left(\frac{\text{Successful Anonymization}}{\text{Total Records}}\right) \times 100$$

This formular calculates the percentage of successfully anonymized records. This metric measures the proportion of records that were successfully anonymized out of the total records.

**Response:** If the percentage of anonymized data falls below a predefined threshold, the system will trigger an alert, prompting the manager to contract the development team to analyze the problem.

4.2. Accuracy (conversion rate)

**Metrics:** Percentage of product recommended items that are purchased at point of sale.

**Measure:** Compare the recommended items against the final purchase. Track the conversion rate of recommendations to actual sales to ensure that system can enhance the business profit.

**Precision@K**

$$\text{Precision@K} = \frac{|\text{Recommended Items @ K} \cap \text{Purchased Items}|}{\text{K}}$$

This formular calculates the precision at a specific rank K, measuring the proportion of relevant recommended items in the top K recommendations. A higher Precision@K indicates better performance, as more of the top recommendations are relevant.

**Recall@K**

$$\text{Recall@K} = \frac{|\text{Recommended Items @ K} \cap \text{Purchased Items}|}{|\text{Purchased Items}|}$$

This formular calculates recall at rank K, measuring the proportion of purchased items that are included in the top K recommended items. A higher Recall@K means that the system is better at suggesting items that users are likely to buy.

**Average Precision@K**

$$\text{Average Precision@K} = \frac{\sum \text{Precision@K (per transaction)}}{\text{Total Transactions}}$$

This formular calculates the average precision across all transactions. This measures how well the recommendation system performs over multiple transactions.

**Average Recall@K**

$$\text{Average Recall@K} = \frac{\sum \text{Recall@K (per transaction)}}{\text{Total Transactions}}$$

This formular calculates the average recall across all transactions. This how well the recommendation system retrieves relevant items over multiple transactions.

**Percentage Meeting Precision Threshold (PMPT)**

$$\text{PMPT} = \left(\frac{\text{Transaction with Precision @K} \geq \text{Threshold}}{\text{Total Transactions}}\right) \times 100$$

This formular calculates the percentage of transactions where Precision@K meets or exceeds the defined threshold. This measure how often the recommendation system achieves a certain level of accuracy, which can be useful in determining the system's reliability.

**Response:** If accuracy drops below a certain level, the system will report accuracy to the manager for further analysis.

4.3. Transparency

**Metrics:** Recommendations with metrics provided.

**Measure:** Implement a feature that logs and how explanations are provided with recommendations (such as confidence, support, lift) to ensure that manager and employee understand the rationale behind the recommendations, track the model result that they always show the metrics with result.

$$\text{Transparency Percentage} = \left(\frac{\text{Valid Metrics}}{\text{Total Model Result}}\right) \times 100$$

This formular calculates the percentage of recommendations that have valid explanations. These metric measures how transparent the model results are by calculating the proportion of valid metrics to the total number of results generated by the model.

**Response:** System will alert prompting the manager to contract the development team to analyze the problem.

4.4. Fairness

**Metric:** Coverage of recommendation among the customers.

**Measure:** Analyze recommendation logs to see how often the system was recommended products to customers.

$$\text{Fairness} = \left(\frac{\text{Transaction with Recommend Items}}{\text{Total Purchased Transactions}}\right) \times 100$$

This formular calculates the percentage of transactions for which both purchased items and recommendations are available. This metric measures how fairly the recommendation system treats all transactions by assessing the proportion of transactions that have received recommendation.

**Response:** If the percentage is too low that shows that system has issue about how it recommends the items. The System notifies the issue to the manager for further consideration.

**5. Key Properties of Recommendations System**

5.1. User Interface

The system will have a user interface to show the model results by listing all the items that they mostly purchase together and show the recommended to additional items for suggestions cashier to inform customers.

5.2. Real-time Data Processing

The system must process transaction data in real-time to provide immediate product recommendations.

5.3. Integration with POS Systems

The system must seamlessly integrate with POS systems to pull the data from POS systems without interrupting the regular sales process.

5.4. Data Storage

The system will extract the items lists in transaction data from POS system and keep them in the recommendation system database.

5.5. Model Update

The system is designed to deploy only on local computers. The system will provide the feature for updating model to let user can retrain model when it degraded.

5.6. Feedback Loop

Allowed user to give the feedback via to developers or support to collect feedback for improvements next version of system.

5.7. Compliance Monitoring

5.7.1. Anonymized percentage for monitoring privacy

5.7.2. Recommendations log for transparency and fairness

5.7.3. Percentage of correction recommendations for accuracy

5.7.4. Fairness of the system.

**6. Data Pipeline and How to Collect Certain Data**

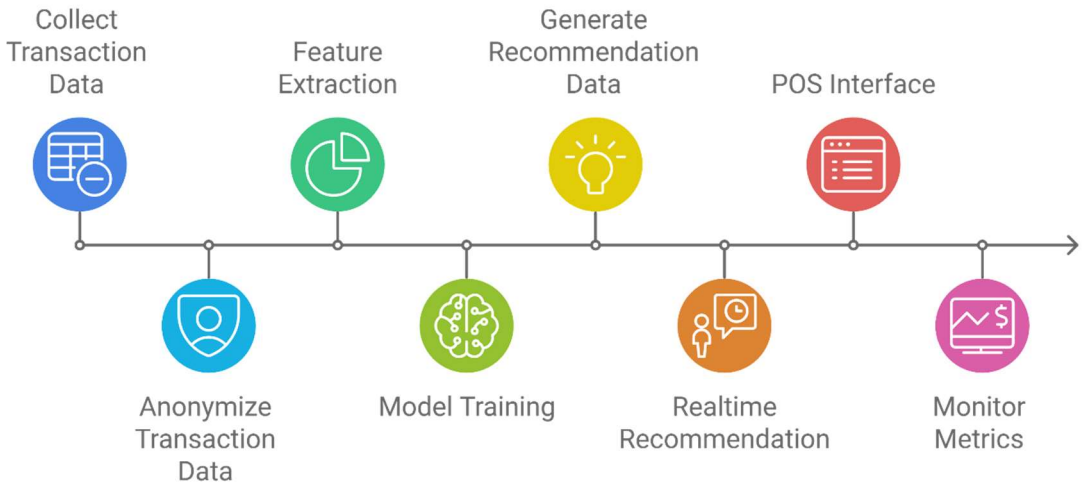


**Figure 1.** Data Pipeline.

6.1. Data Collection

**Source:** The primary data will come from POS systems within the stores. This data includes the details of transactions record such as products purchase, date of transactions was created, customer ID, price, etc.

**Process:** Data is transmitted to the recommendation system when the transaction is completed and preprocessed before saving to the system database.

6.2. Data Preprocessing (Anonymize transaction data)

**Anonymize and Indexing:** add feature `Transaction_id` to each record of data to use it as index, by this Transaction feature not related to Transaction in POS system. And create log table to ensure that the complete anonymization was recorded.

**Cleaning:** Check if data has missing values and drop them to ensure data quality.

**Data Storage:** All the data will be stored in recommendation database which consist of four tables which are Transaction to store transaction products that were anonymized, Anonymization logs to record status of anonymization, Association rules to store the model results, Recommendation logs to store recommendations in transaction. The schema of tables is shown in following

**Table 1.** Transaction table schema.

| Column names | Data type | Description | constrain |
|---|---|---|---|
| Transaction_id | PRIMARY KEY | Index the transactions. | AUTO INCREATMENT |
| Products | TEXT | Collect list of products | - |
| Datetime | TEXT | Date that was collected | - |

**Table 2.** Anonymization logs table schema.

| Column names | Data type | Description | constrain |
|---|---|---|---|
| Transaction_id | PRIMARY KEY | Index the transactions. | AUTO INCREATMENT |
| Anonymization_Timestamp | TEXT | Date that the data was anonymized. | - |
| Status | TEXT | Status of anonymization there are Success/Failed (e). | - |

**Table 3.** Association rules table schema.

| Column names | Data type | Description | constrain |
|---|---|---|---|
| Antecedents | TEXT | The product that was chosen first. | - |
| Consequents | TEXT | The product that mostly was chosen with the first product. | - |
| Support | REAL | Represents the proportion of transactions that contain both the antecedent and the consequent. | - |

| | | | |
|---|---|---|---|
| Confidence | REAL | Indicates the likelihood of the consequent being purchased if the antecedent is purchased. | - |
| Lift | REAL | Measures how much more likely the consequent is bought when the antecedent is bought, compared to its usual frequency. | - |
| Leverage | REAL | Represents the difference between the observed co-occurrence of antecedent and consequent and the expected co-occurrence if they were independent. | - |

**Table 4.** Recommendation table schema.

| Column names | Data type | Description | constrain |
|---|---|---|---|
| Transaction_id | TEXT | Index the transactions. | - |
| Recommended_items | TEXT | Collect list of products. | - |
| Purchased_items | TEXT | Date that was collected. | - |
| timestamp | TEXT | The date that transaction was create. | - |

6.3. Feature Extracting

The system begins with the preprocessing and training of historical data. In this project, use the data from table transaction to be the foundation for the initial model training. Firstly, the dataset is cleaned by removing missing values and duplicate rows. The cleaned data is transformed into a suitable format using the Transaction Encoder.

6.4. Model Training

After the dataset has been preprocessed, FP-Growth Algorithm is employed with the minimum support 3.5%, minimum lift 1.5 and minimum confidence 80% (can vary depend on

stakeholder and data character of each store) to ensure the generation of high-quality recommendations.

The FP Growth algorithm refers to frequent patterns. Support is the percentage of transactions in the dataset that contain a particular item or set of items. Which can calculate by following formular.

$$\text{Support (item lists)} = \frac{\text{Number of transactions containing (item list)}}{\text{Total Number of Transactions}}$$

Confidence indicates the possibility that a product will be bought given another item is purchased. Which can calculate by following formular.

$$\text{Confidence (antecedents} \rightarrow \text{consequents)} = \frac{\text{Support (antecedents} \cup \text{consequents)}}{\text{Support (antecedents )}}$$

Lift measures the strength and significant association between two items. item is purchased. Which can calculate by following formular.

$$\text{Lift (antecedents} \rightarrow \text{consequents)} = \frac{\text{Support (antecedents} \cup \text{consequents)}}{\text{Support (antecedents )} \times \text{Support (consequents)}}$$

Leverage measures the difference between the observed co-occurrence of two items and what would be expected if they were independent. Which can calculate by following formular.

$$\text{Leverage (antecedents} \rightarrow \text{consequents)} = \text{Support (antecedents} \cup \text{consequents)} - \text{Support (antecedents)} \times \text{Support (consequents)}$$

The model results generated associations rules and stores them in system database. This ensures transparency in the recommendation process.

In case that model is degraded system have feature to retain model due to the customer buying behavior has changed. The model will reprocess since the collection of new data from POS and complete retraining process reflects current purchasing patterns and emerging trends ensuring the system stays aligned with customer behavior.

## 7. System Design

Figure 2 depicts a recommendation system architecture that integrates with a point of sale (POS) system. The role of each component in the system:

### 7.1. POS System

The POS system collects real-time transaction data from the store. Each customer transaction is sent to the system, which includes details about the products being purchased, location, customer information.

### 7.2. Anonymized Transaction Data

The raw transaction data from the POS system is anonymized before it saves in the recommendation database. This step ensures that sensitive information about customers is

removed in compliance with privacy regulations such as the PDPA (Personal Data Protection Act) and generates log save in database.

7.3. Data Storage

The anonymized transaction data is stored in a central database. This database serves as the foundation for both model training and generating real-time recommendations and save log for anonymization and recommendations.

7.4. Model Training

The stored transaction data is used to train machine learning models. These models analyze patterns in customer purchases, identifying relationships between products that are frequently bought together. The results of this model training are stored in a database for later use.

7.5. Model Results

The results of the trained models are stored in the system. This includes association rules that show the insights derived from the data, which will be used to recommend. These recommendations could help with product placement, cross-selling strategies, or promotions.

7.6. Recommendation Generation

Based on the model result and current customer transaction data, the recommendation module generates product suggestions for the real time recommendation.

7.7. Metrics Monitoring

The performance of the recommendation system is continuously monitored through metrics like Anonymized percentage, Precision@K, Recall@K, ...etc. Monitoring helps ensure the recommendations are accurate and effective, allowing for improvements or adjustments to the model if necessary and ensure that the systems compliance with Thailand act.

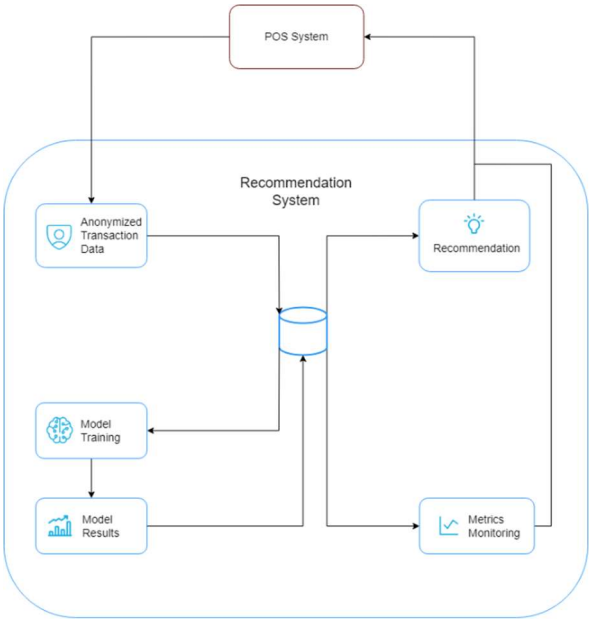


**Figure 2.** System design overview.

## 8. Human in the loop

The recommendations from the system will have human from stakeholder including store manager, cashier to decision-makers the output both in real time and shelf recommendation. And they also need to evaluate the metrics monitoring to track if the system fails to deliver appropriate suggestions. The store employee should provide feedback on the accuracy and effectiveness of the system to the developer team.

## 9. Compliance with Thailand regulations

Since our system is intended to deploy in Thailand, we have ensured the system continues to comply with Thailand's AI Policy Framework. All customer data is anonymized before processing. Conversion rate of recommended products is tracked at the POS system, ensuring that recommendations work. We used FP growth algorithm, which is interpretable, for model training. No sensitive data like genders and race, are used to train the model ensuring that the model will not make any discrimination based on this content when it makes recommendation.

9.1. Privacy

The recommendation system anonymously all transactions that were fetch to system by using only products lists in transactions. Also provide log to record that data were successfully or failed for anonymous transactions.

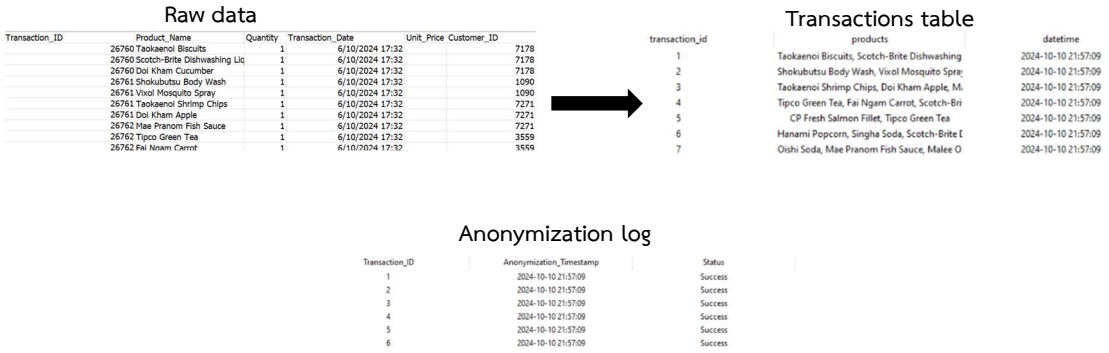


**Figure 3.** Privacy compliance.

9.2. Security

The recommendation system will be deployed in offline environments which the person that can access the data is the manager and employees of those stores. Also implement authentication to limit access to model results and logs.

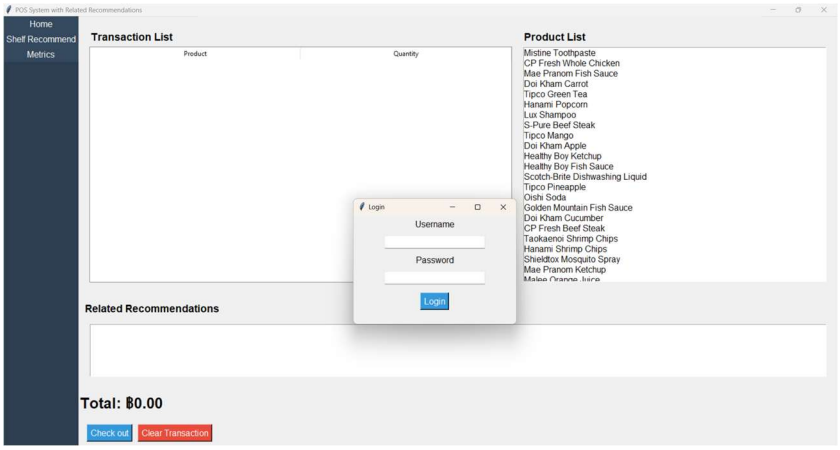


**Figure 4.** Security compliance.

9.3. Reliability

The system continues to track the accuracy (Precision@K, Recall@K, and PMPT metrics) of the model and gives warning if model accurate fell below the system threshold.

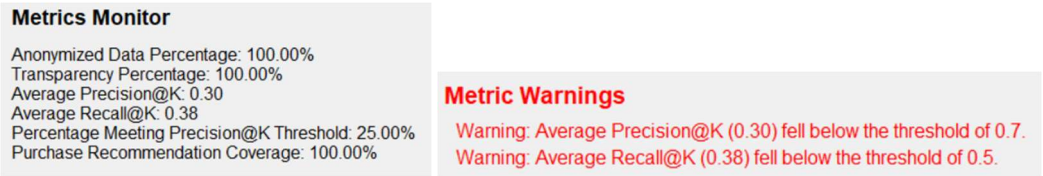


**Figure 5.** Reliability compliance.

9.4. Fairness

The system tracks and logs the recommendation for each complete transaction to ensure that all customers receive recommendations.

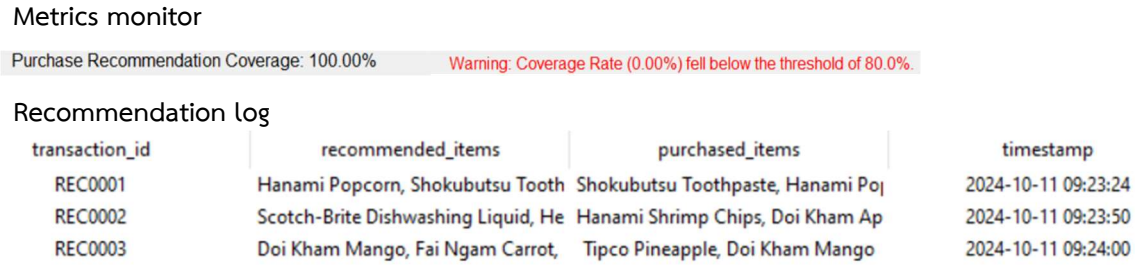


**Figure 6.** Fairness compliance.

9.5. Transparency

The system always provides explainability features including logs showing support, confidence, lift and leverage for each recommendation.

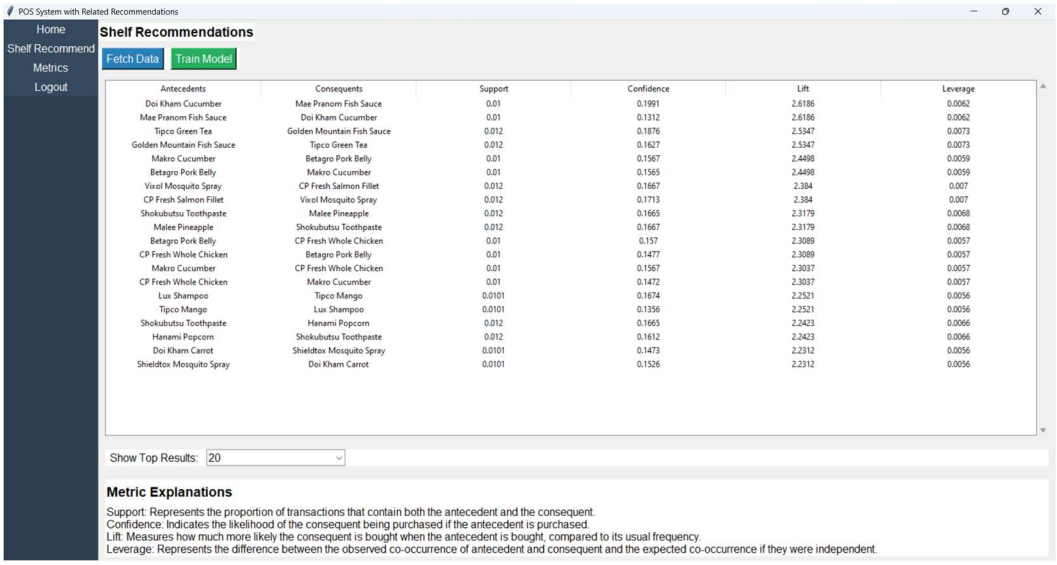


**Figure 7.** Transparency compliance.

9.6. Accountability

The system always calls humans when it needs human intervention with the system by giving function to send feedback to developer or retrain model.

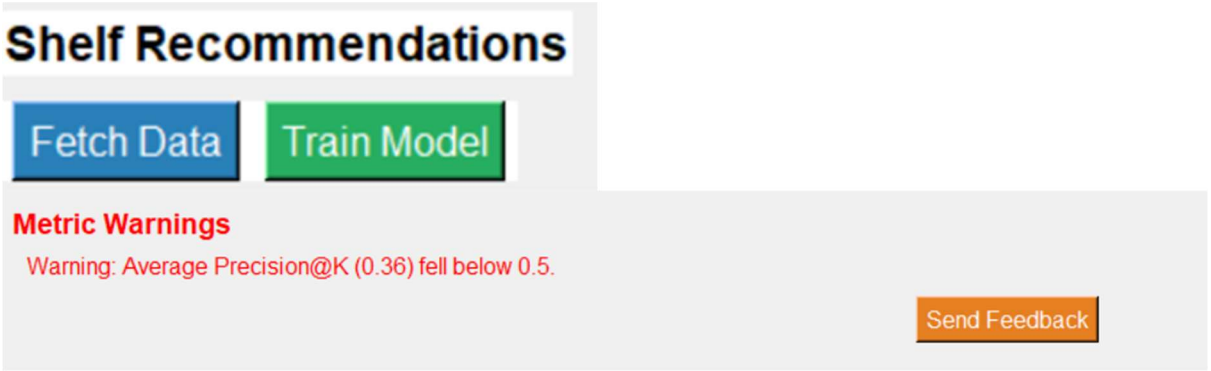


**Figure 8.** Accountability compliance.

9.7. Human Oversight and Human Agency

The utilization of model recommendations is based on manager and employee decision-making processes. Which The system will not make decisions process if there is no human intervention.

## 10. Conclusion

The development and implementation of an AI-based recommendation system designed specifically for local convenience stores in Thailand. The system enhances retail operations by analyzing historical transaction data to recommend product groupings and optimize shelf arrangements, helping to drive sales and improve customer experience.

The system is designed with key dimensions aligned with Thailand's AI Policy Framework, including Privacy, Security, Reliability, Fairness, Transparency, Accountability, and Human-Centricity. It ensures customer data is anonymized, the system is secure, and recommendations are explainable and reliable. The system is primarily used by store managers and employees, maintaining human oversight in decision-making processes.

Several metrics were defined to monitor system performance, including Anonymized Percentage, Precision@K, Recall@K, and PMPT (Percentage Meeting Precision Threshold), transparency percentage, Recommendation coverage. These metrics track the effectiveness and compliance of the system in real-time and provide feedback to help improve the system's accuracy, fairness, and transparency over time.

Integration with Point-of-Sale (POS) systems allows for seamless data collection, model retraining, and feedback loops to continuously align the recommendations with evolving customer behaviors. The system is designed to comply with Thailand's Personal Data Protection Act (PDPA) and other relevant regulations, ensuring legal adherence in its deployment.

**Essential link**

**Project source code:** https://github.com/KaySokay/Recommendation_System_SC348810