

Milestone 1 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

Milestone 1

Declaration: Generative AI is used to translate documents and extract knowledge from documents.

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-driven 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 items 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 items 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 items 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.

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 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 and FP-Growth algorithms which is in 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. Metric 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.

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.

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 an explanation 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.

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

4.4. Fairness

Metric: Distribution of recommended products.

Measure: Analyze recommendation logs to see the distribution of recommended products.

Response: If the distribution of recommendations shows significant bias. System will report the issue to the manager for consideration about issue.

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 and keep in csv file to prepare for training the model.

5.5. Model Update

The system is designed to deploy only on local computers which means the model will retrain itself every time when reopen system.

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

6. Data Pipeline and How to Collect Certain Data

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 purchase in the following format

Table 1. Transaction format for collection.

List of Products	Datetime
(Product 1, Product 2, ..., Product n)	DD-MM-YYYY Time

Note: n is the number of products; D, M, Y are the abbreviations of Day, Month, Year respectively.

Process: Data is transmitted to the recommendation system in real time and then data will be saved to CSV file when the transaction is completed.

6.2.Data Preprocessing

Anonymize and Indexing: add feature 'Transaction' to each record of data to use it as index, by this Transaction feature not related to Transaction in POS system. And create logs file to ensure that anonymized is complete.

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

6.3.Data Storage

The data for AI training will be stored in CSV file by following these naming and structure for storage data to train the model.

CSV Naming:

Transaction file when each transaction from POS system is done, the system will collect data and save it to CSV file in name format 'transaction_YYYY-MM-DD.csv' for easy to manage.

Log file will save transactions log when it successfully.

CSV Structure:

In transaction CSV file will has

`transaction`: The number from 1 to n.

`products`: The list of products from each transaction in POS system will be in format such as apple, banana, chocolate, etc.

`datetime`: The date and time of the transaction.

In log CSV file will has

`transaction_index`: The number of transactions index

`anonymization_timestamp`: The date and time of the transaction.

`status`: Status of saving transactions

Table 2. Transaction data in CSV file.

Transaction	List of Products	Datetime
X	(Product 1, Product 2, ..., Product n)	DD-MM-YYYY Time

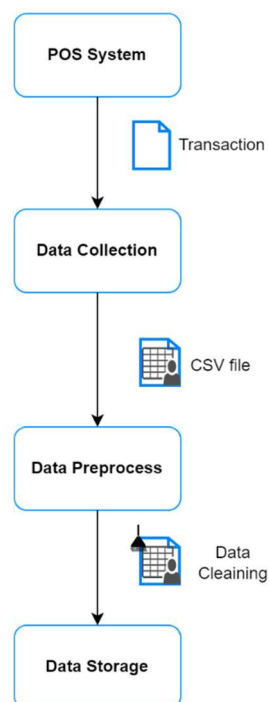
Note: X is the number of rows to Xn transactions; n is the number of products; D, M, Y are the abbreviations of Day, Month, Year respectively.

Table 3. Log data in CSV file.

Transaction	Anonymization_Time	Status
X	DD-MM-YYYY Time	True/False

Note: X is the number of rows to Xn transactions; D, M, Y are the abbreviations of Day, Month, Year respectively.

Figure 1. Data Pipeline



Essential link

Data pipeline code: https://github.com/KaySokay/Recommendation_System_SC348810

Example Dataset: <https://pbpython.com/market-basket-analysis.html>