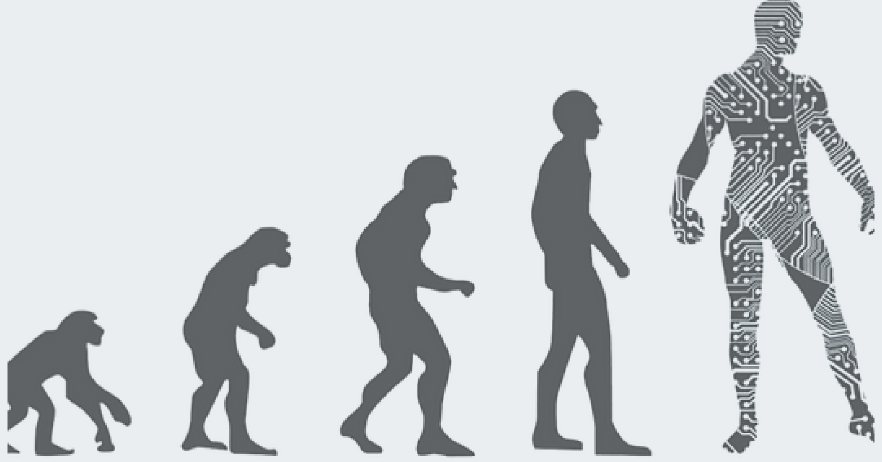
RECONNENDATION AI-ENABLED SYSTEM FOR LOCAL STORE

Software Development and Project Management Final Poject Presentation

675380013-9 Somkiad Khenwiset
675380020-2 Thant Lwin Oo

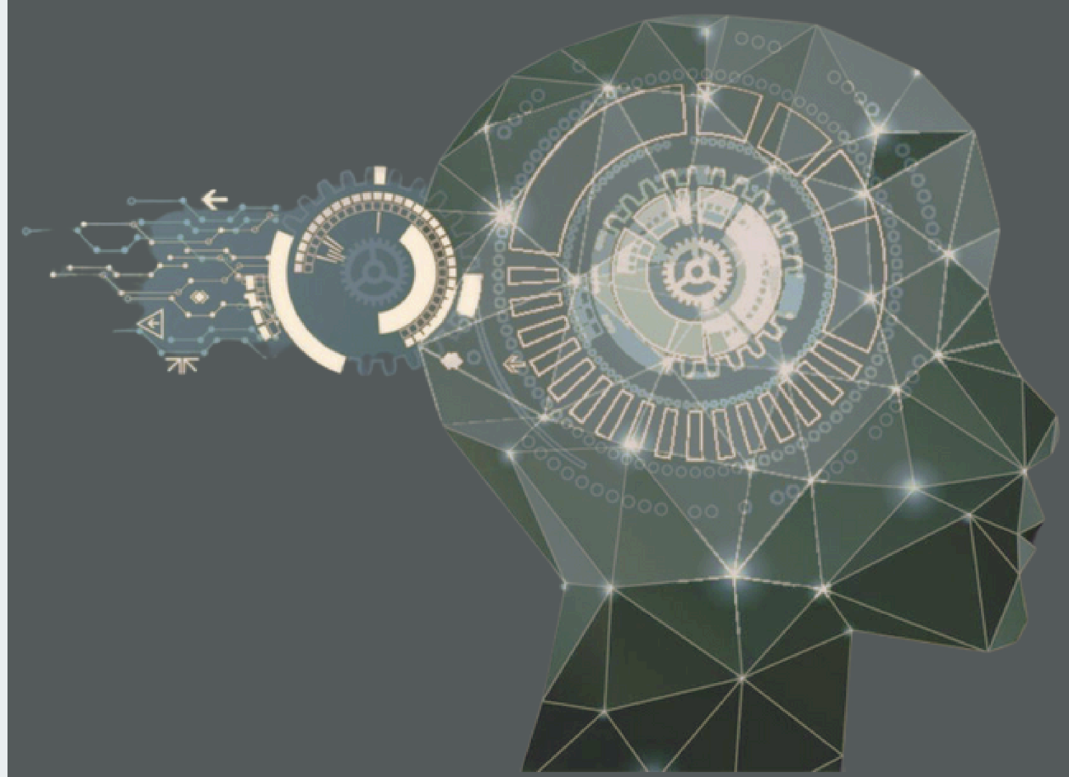
INTRODUCTION



Today's, as the Evolution of AI technology reaches maturity, it is widely leveraged across various sectors, without even noticing.

This transformation significantly impacts our daily lives and business operations.

Today, we are launching a project that combines **AI and Marketing** for innovative solutions.



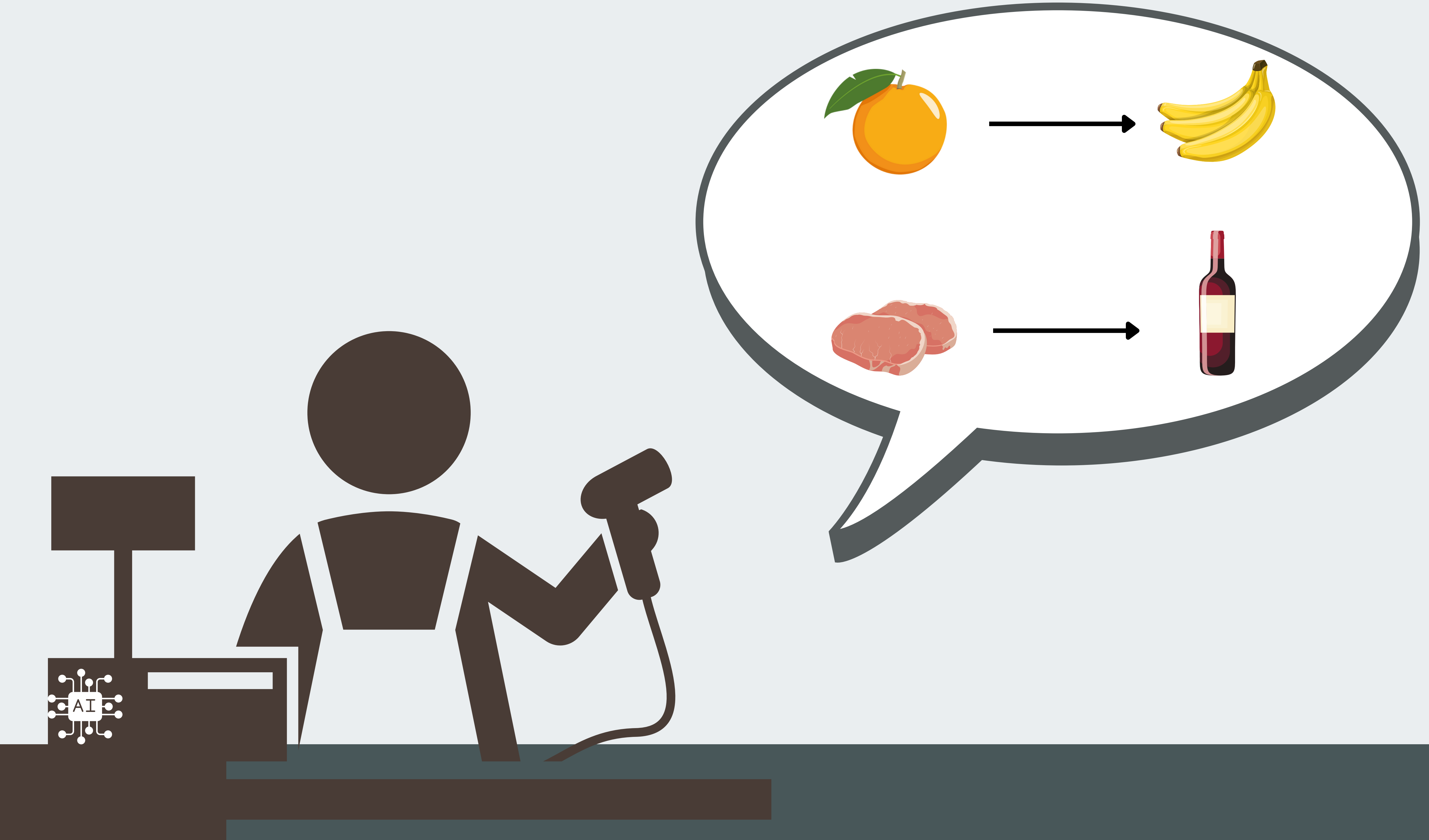
WHERE TO LUNCH?

Thailand

PROBLEM DOMAIN

"Imagine walking into a store and effortlessly finding items you didn't know you needed—how would that impact your shopping experience?"

AI-BASED RECOMMENDATION FOR LOCAL STORES



AI-BASED RECOMMENDATION FOR LOCAL STORES: EXPLAIN

The AI-driven recommendation system analyzes historical purchase data to identify product patterns frequently bought together It provides recommendations to optimize shelf arrangements and real-time helps cashiers suggest complementary items at checkout. This system improves the shopping experience and boosts sales by promoting additional products customers may not have initially considered.

THAILAND AI POLICY FRAMEWORK

Privacy

AI system must protect personal data by ensuring no misuse of sensitive information. Compliance with Thailand’s PDPA requires explicit consent for data collection, data anonymization, and secure management of storage and access.

Security

AI systems must implement strong security measures to prevent data breaches and unauthorized access. This includes data encryption, regular security audits, access controls, and mechanisms for quick detection and response to security incidents.

Reliability

AI systems must ensure trust by delivering accurate, reliable, and reproducible results. Maintaining data quality control is essential to prevent errors and protect credibility, especially given AI's uncertain impacts.

THAILAND AI POLICY FRAMEWORK

Fairness

AI systems should be designed to promote fairness, equality, and inclusivity, ensuring all societal groups, especially the disadvantaged, benefit equally, without bias or discrimination.

Transparency

AI systems should be designed for human understanding and oversight, ensuring data usage and decisions are explainable, traceable, and auditable to build trust and accountability. AI capabilities and limitations should be communicated clearly and appropriately, tailored to the audience's expertise.

Accountability

AI systems should ensure accountability and clear traceability for any impacts, with provisions to address potential damages. All stakeholders should consult and plan for managing long-term risks and impacts.

Human-Centricity

AI systems should prioritize a human-centric approach, keeping humans in control of critical decisions. They should be designed to benefit humanity, avoid harm, uphold human values, and support sustainable development for society and the environment.

STAKEHOLDERS

- Store Owners
- Store Employees
- POS System Providers
- Developers
- Customer

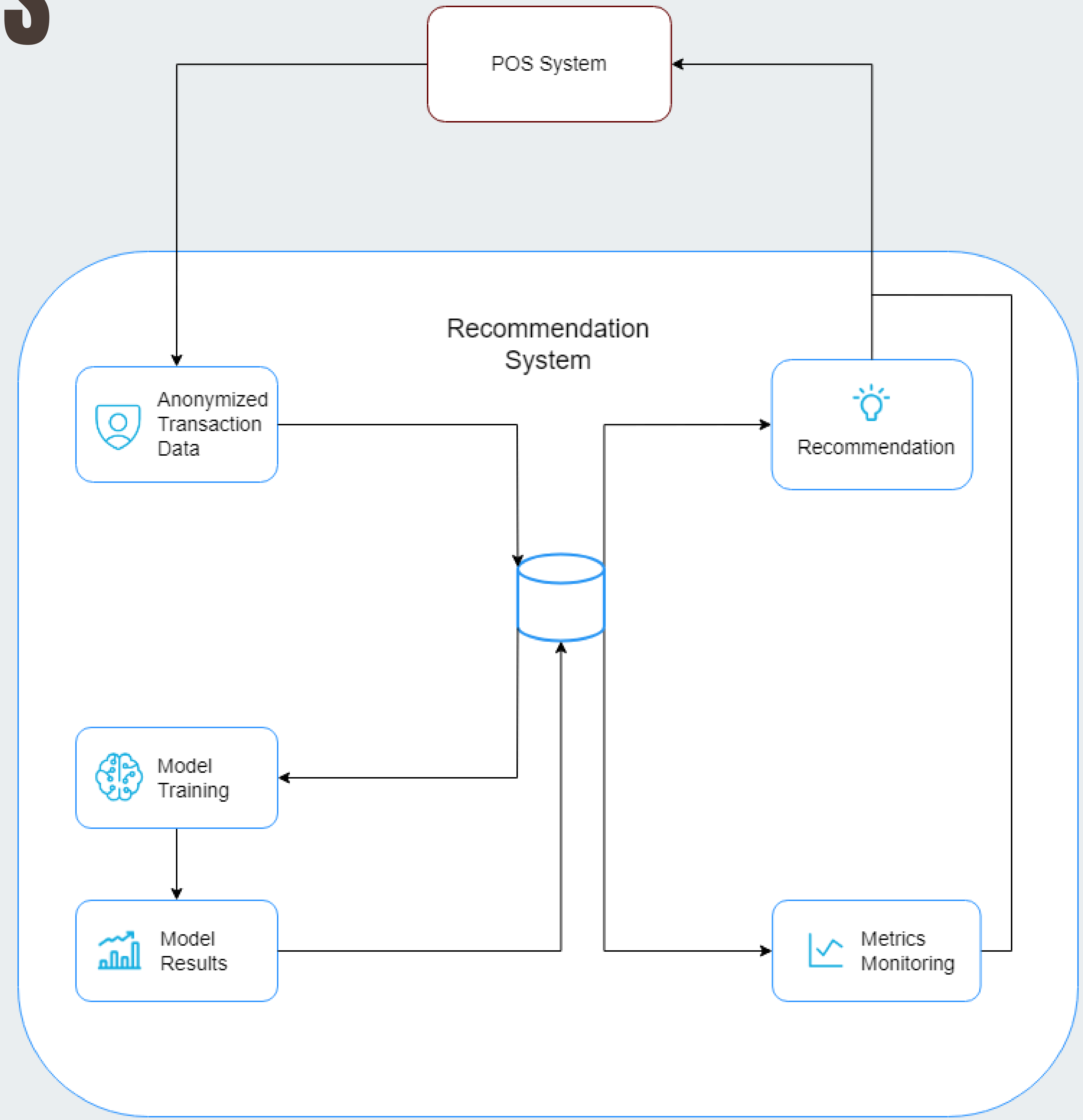
SYSTEM OVERVIEWS

Python 3.12

MLXTEND 0.20

matplotlib 3.6.0

pandas 1.5.0



ANONYMIZATION

FUNCTION anonymize_data(data):
  FOR each transaction IN data:
    REMOVE not relevant data (e.g., Customer_ID)
    LOG anonymization status (Success/Failure)
  RETURN anonymized data

Transaction_ID	Product_Name	Quantity	Transaction_Date	Unit_Price	Customer_ID
1	Taokaenoi Biscuits	1	6/10/2024 17:32		7178
1	Scotch-Brite Dishwashing	1	6/10/2024 17:32		7178
1	Doi Kham Cucumber	1	6/10/2024 17:32		7178
2	Shokubutsu Body Wash	1	6/10/2024 17:32		1090
2	Vixol Mosquito Spray	1	6/10/2024 17:32		1090

transaction_id	products	datetime
1	Taokaenoi Biscuits, Scotch-Brite Dishwashing	2024-10-11 09:22:48
2	Shokubutsu Body Wash, Vixol Mosquito Spray	2024-10-11 09:22:48
3	Taokaenoi Shrimp Chips, Doi Kham Apple, Ma	2024-10-11 09:22:48
4	Tipco Green Tea, Fai Ngam Carrot, Scotch-Bri	2024-10-11 09:22:48

FUNCTION save_anonymized_transactions(anonymized data):
  FOR each transaction IN data:
    COMBINE products into a single string
    INSERT products and timestamp
    INTO transactions table
    INSERT anonymization log (timestamp, status)
    INTO anonymization_logs table
  COMMIT transaction logs to database

Transaction_ID	Anonymization_Timestamp	Status
1	2024-10-11 09:22:48	Success
2	2024-10-11 09:22:48	Success
3	2024-10-11 09:22:48	Success
4	2024-10-11 09:22:48	Success
5	2024-10-11 09:22:48	Success
6	2024-10-11 09:22:48	Success

DATA PREPROCESSING

- **Data Collection**

Transaction data from POS system is anonymized and stored in the storage.

For this project we will use an exmple dataset to train the model.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	1/12/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	1/12/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	1/12/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	1/12/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	1/12/2010 8:26	3.39	17850.0	United Kingdom

- **Data Prepreparation**

After dropping null and duplicate values, the dataset is group by Invoice number and Description. No sensitive column names likes genders and race, are used to train the models. This ensures that the model will not make any discrimination based on this contents when it makes recommendation.

	InvoiceNo	Products
0	536365	[WHITE HANGING HEART T-LIGHT HOLDER, WHITE MET...
1	536366	[HAND WARMER UNION JACK, HAND WARMER RED POLKA...
2	536367	[ASSORTED COLOUR BIRD ORNAMENT, POPPY'S PLAYHO...
3	536368	[JAM MAKING SET WITH JARS, RED COAT RACK PARIS...
4	536369	[BATH BUILDING BLOCK WORD]

MODEL TRAINING

For model training, **FP-Growth Algorithm,** which compliances the transparency, is used to train the model.

Why We Choose FP Growth

FP-Growth is more efficient than the Apriori algorithm, it compresses the dataset into a compact FP-tree (Frequent Pattern tree) and mines patterns directly from the tree, reducing the number of scans times to dataset.

Since FP-Growth reduces the need for multiple database scans, it can efficiently handle very large transaction datasets without overwhelming computational resources.

ECLAT is another algorithm used for frequent pattern mining, FP-Growth can handle long frequent itemsets more effectively, as the tree structure compresses repeated patterns into a compact representation.

FP-GROWTH ALGORITHM

The FP-Growth algorithm refers to **frequent patterns**.

Support

Support is the percentage of transactions in the dataset that contain a particular item or set of items.

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

Confidence

Confidence indicates the possibility that a product will be bought given another item is purchase.

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

Confidence

Lift measures the strength and significant association between two items.

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

WORK FLOW OF FP-GROWTH

To understand more about some metrics that are used in model training and inner workings of FP Growth algorithm, we will explain with a small dataset.

This will ensure that there is no blackbox in the model and inhance to solve the transparency issue.

Transaction_Id	Items
1	Milk
2	Milk, bread, Saffron
3	Milk,bread
4	Cold Drink, Wafer
5	Cold Drink
6	Cold Drink, Milk, Wafer

$$\text{Support (milk)} = \frac{\text{No of Times Milk was purchased}}{\text{Total Transaction}}$$

$$= \frac{4}{6}$$

$$\text{Support (bread)} = \frac{2}{6}$$

$$\text{Support (Milk, bread)} = \frac{2}{6}$$

$$\text{Confidence (milk, bread)} = \frac{\text{No of Times Milk and bread was purchased together}}{\text{No of Times milk was purshased}}$$

$$= \frac{2}{4}$$

$$\text{Confidence (bread,milk)} = \frac{\text{No of Times Milk and bread was purchased together}}{\text{No of Times bread was purshased}}$$

$$= \frac{2}{2}$$

$$\text{lift (milk, bread)} = \frac{\text{Support (Milk,Bread)}}{\text{support (Milk) * support(Bread)}}$$

WORK FLOW OF FP GROWTH

Frequency Counting

First of all, we need to count how many times did the items are frequently appear in our dataset.

Transaction_Id	Items
1	Milk
2	Milk, bread, Saffron
3	Milk,bread
4	Cold Drink, Wafer
5	Cold Drink
6	Cold Drink, Milk, Wafer

Items	Frequency
Milk	4
bread	2
saffron	1
Cold Drink	3
Wafer	2

WORK FLOW OF FP-GROWTH

Defining Minimum Support Rearrange the Items

Let’s think about that, we defined as minimum support is 2 in this expamle. We will drop the items which is less than 2 times. After that, the items are reordered for each transaction according to the frequency.

Transaction_Id	Items
1	Milk
2	Milk, bread, Saffron
3	Milk,bread
4	Cold Drink, Wafer
5	Cold Drink
6	Cold Drink, Milk, Wafer

Items	Frequency
Milk	4
bread	2
saffron	1
Cold Drink	3
Wafer	2

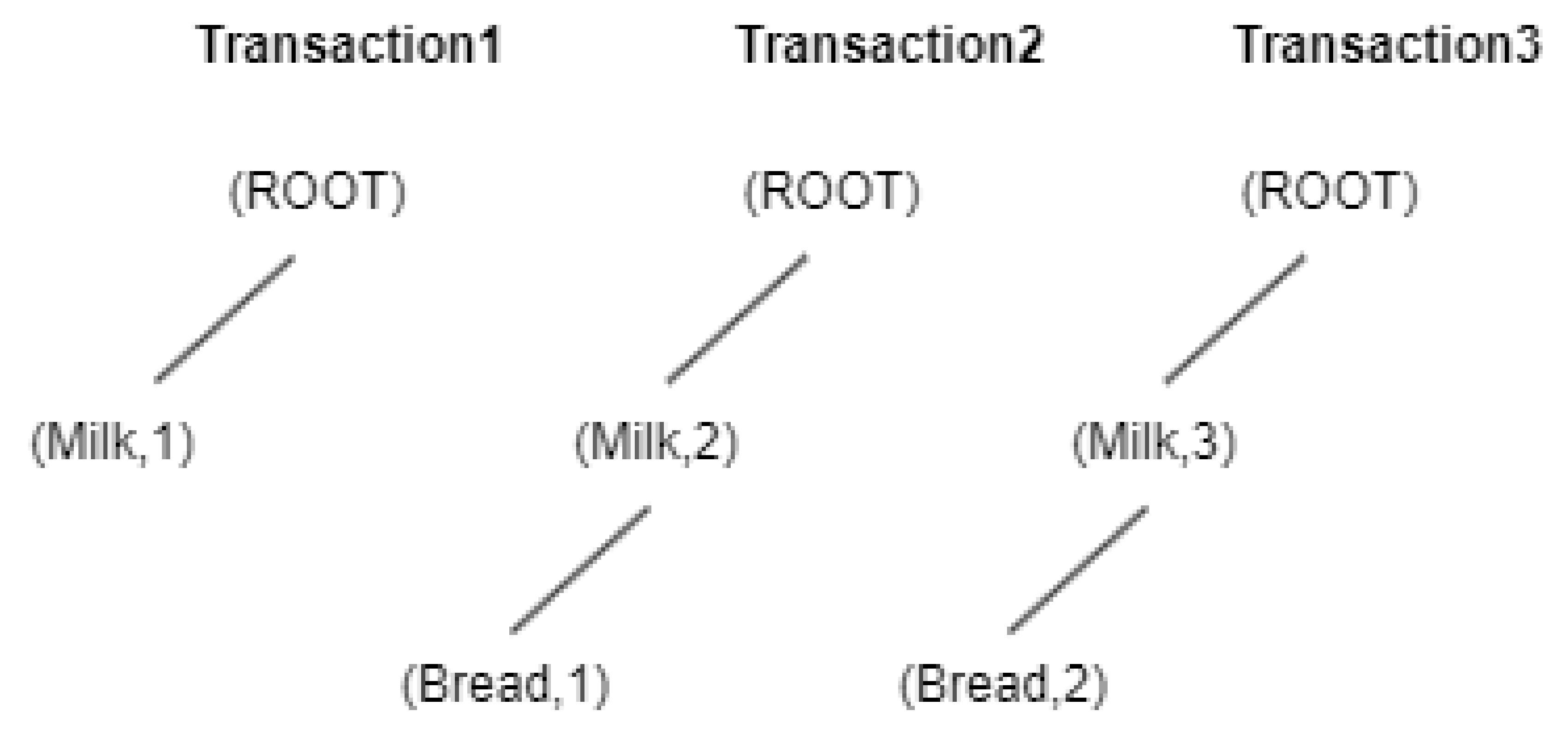
Transaction_Id	Items	Rearrange Items
1	Milk	Milk
2	Milk, bread, Saffron	Milk, bread
3	Milk,bread	Milk, bread
4	Cold Drink, Wafer	Cold Drink, Wafer
5	Cold Drink	Cold Drink
6	Cold Drink, Milk, Wafer	Milk, Cold Drink, Wafer

WORK FLOW OF FP-GROWTH

FP-Tree Construction

Based on the rearranged transation, FP-Tree is built step by step.

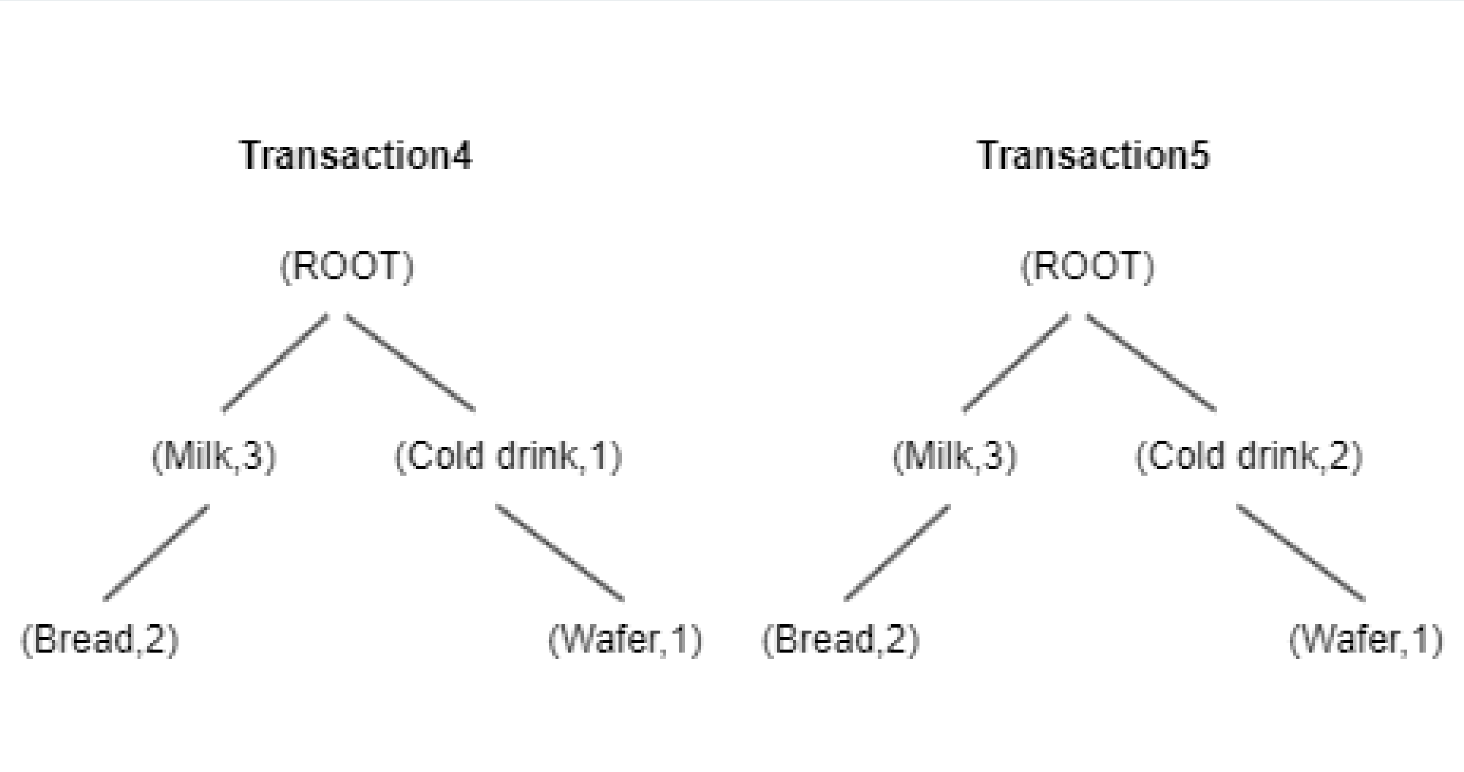
Transaction_Id	Items	Rearrange Items
1	Milk	Milk
2	Milk, bread, Saffron	Milk, bread
3	Milk,bread	Milk, bread
4	Cold Drink, Wafer	Cold Drink, Wafer
5	Cold Drink	Cold Drink
6	Cold Drink, Milk, Wafer	Milk, Cold Drink, Wafer

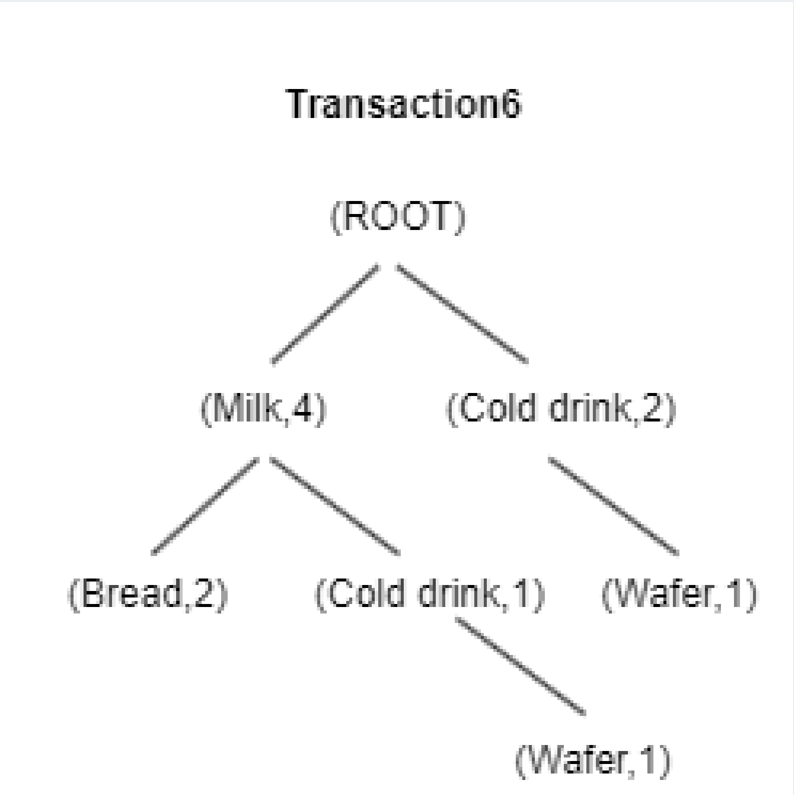


WORK FLOW OF FP-GROWTH

FP-Tree Construction

Transaction_Id	Items	Rearrange Items
1	Milk	Milk
2	Milk, bread, Saffron	Milk, bread
3	Milk,bread	Milk, bread
4	Cold Drink, Wafer	Cold Drink, Wafer
5	Cold Drink	Cold Drink
6	Cold Drink, Milk, Wafer	Milk, Cold Drink, Wafer



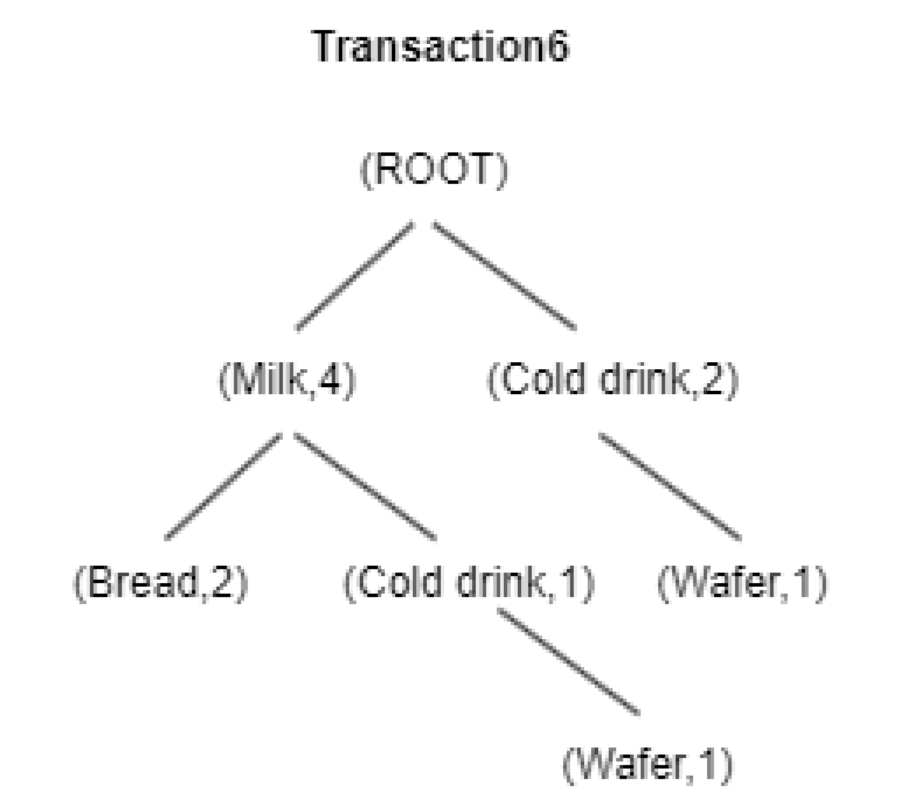


WORK FLOW OF FP-GROWTH

Frequent Item set Generation

Next Step is generation of frequent items from the last FP-Tree.

Transaction_Id	Items	Rearrange Items
1	Milk	Milk
2	Milk, bread, Saffron	Milk, bread
3	Milk,bread	Milk, bread
4	Cold Drink, Wafer	Cold Drink, Wafer
5	Cold Drink	Cold Drink
6	Cold Drink, Milk, Wafer	Milk, Cold Drink, Wafer



Item	Conditional Pattern Base	Support Count	Conditional FP Tree (Min Support = 2)	Frequent Itemset
(Bread,2)	(Milk,2)	(Milk,2)	(Milk,2)	Milk,bread
(Wafer,2)	(Cold Drink,1) (Cold Drink,1 : Milk1)	(Cold Drink,2 : Milk1)	(Cold Drink,2)	Cold Drink, Wafer
(Cold Drink,3)	(Milk,1)	(Milk,1)	-	-
(Milk,4)	-	-	-	-

WORK FLOW OF FP-GROWTH

Generation Rules betwen Frequent Itemsts

In this case, we have two frequent item sets: {milk, bread} {Cold Drink, Wafer}. So there will be four possibilites as shown in table.
To generate the recommandations from the system, we need to identify minimum cofident and minimum lift.
Let’s think a scenarion that the minimum confident => 0.8 and minimum lift is 1.1.

antecedent	consequent	Confident	Lift
Milk	Bread	0.8	1.2
Bread	Milk	0.7	1.2
Cold Drink	Wafer	0.8	1.6
Wafer	Cold Drink	0.7	1.6

The system will only recommend that a person who buys milk should also buy bread, but it will not recommend buying milk for a person who buys bread.

MODEL TRAINING

Ok, Now Let's go back to Model Training Process for our Project.

In the project, we defined **minimum support = 0.035**.

It means that if we have 1000 transactions, the items need to appear in a least 35 transaction to be considerd as frequent.

Confidence thredshold is **0.8,** and **Lift threadshold** is **1.5**.

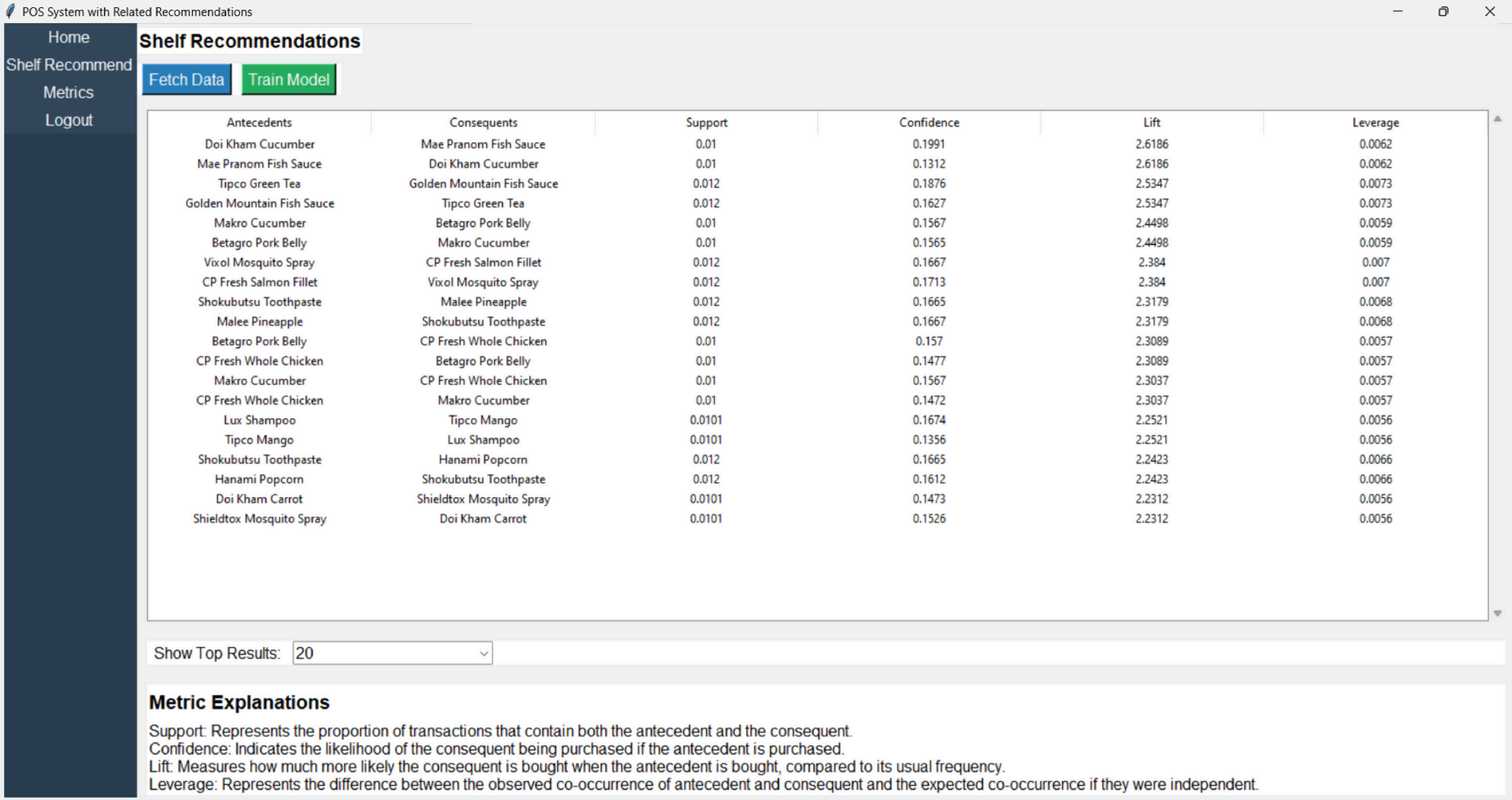
The model produces recommendations only for items that have at least 80% confidence and a lift value of 1.5.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
frozenset({"POPPY'S PLAYHOUSE LIVINGROOM "})	frozenset({"POPPY'S PLAYHOUSE BEDROOM "})	0.036281179138322000	0.05215419501133790	0.036281179138322000	1.0	19.17391304347830
frozenset({"POPPY'S PLAYHOUSE KITCHEN"})	frozenset({"POPPY'S PLAYHOUSE BEDROOM "})	0.05215419501133790	0.05215419501133790	0.045351473922902500	0.8695652173913040	16.67296786389410
frozenset({"POPPY'S PLAYHOUSE BEDROOM "})	frozenset({"POPPY'S PLAYHOUSE KITCHEN"})	0.05215419501133790	0.05215419501133790	0.045351473922902500	0.8695652173913040	16.67296786389410
frozenset({'ALARM CLOCK BAKELIKE RED '})	frozenset({'ALARM CLOCK BAKELIKE GREEN'})	0.047619047619047600	0.05668934240362810	0.04081632653061220	0.8571428571428570	15.12
frozenset({'RED WOOLLY HOTTIE WHITE HEART.', 'RETRO COFFEE MUGS ASSORTED'})	frozenset({'KNITTED UNION FLAG HOT WATER BOTTLE'})	0.036281179138322000	0.08390022675736960	0.036281179138322000	1.0	11.918918918918900
frozenset({'KNITTED UNION FLAG HOT WATER BOTTLE', 'RETRO COFFEE MUGS ASSORTED'})	frozenset({'RED WOOLLY HOTTIE WHITE HEART.'})	0.03854875283446710	0.08616780045351470	0.036281179138322000	0.9411764705882350	10.922600619195000
frozenset({'RED WOOLLY HOTTIE WHITE HEART.', 'WHITE HANGING HEART T-LIGHT HOLDER'})	frozenset({'KNITTED UNION FLAG HOT WATER BOTTLE'})	0.047619047619047600	0.08390022675736960	0.04308390022675740	0.9047619047619050	10.783783783783800

These are some example results from our model and the recommendation and shelf arrangent will be considered based on these relevant rules.

RECOMMENDATION

Shelf Recommendation



Antecedents	Consequents	Support	Confidence	Lift	Leverage
Doi Kham Cucumber	Mae Pranom Fish Sauce	0.01	0.1991	2.6186	0.0062
Mae Pranom Fish Sauce	Doi Kham Cucumber	0.01	0.1312	2.6186	0.0062
Tipco Green Tea	Golden Mountain Fish Sauce	0.012	0.1876	2.5347	0.0073
Golden Mountain Fish Sauce	Tipco Green Tea	0.012	0.1627	2.5347	0.0073
Makro Cucumber	Betagro Pork Belly	0.01	0.1567	2.4498	0.0059
Betagro Pork Belly	Makro Cucumber	0.01	0.1565	2.4498	0.0059
Vixol Mosquito Spray	CP Fresh Salmon Fillet	0.012	0.1667	2.384	0.007
CP Fresh Salmon Fillet	Vixol Mosquito Spray	0.012	0.1713	2.384	0.007
Shokubutsu Toothpaste	Malee Pineapple	0.012	0.1665	2.3179	0.0068
Malee Pineapple	Shokubutsu Toothpaste	0.012	0.1667	2.3179	0.0068
Betagro Pork Belly	CP Fresh Whole Chicken	0.01	0.157	2.3089	0.0057
CP Fresh Whole Chicken	Betagro Pork Belly	0.01	0.1477	2.3089	0.0057
Makro Cucumber	CP Fresh Whole Chicken	0.01	0.1567	2.3037	0.0057
CP Fresh Whole Chicken	Makro Cucumber	0.01	0.1472	2.3037	0.0057
Lux Shampoo	Tipco Mango	0.0101	0.1674	2.2521	0.0056
Tipco Mango	Lux Shampoo	0.0101	0.1356	2.2521	0.0056
Shokubutsu Toothpaste	Hanami Popcorn	0.012	0.1665	2.2423	0.0066
Hanami Popcorn	Shokubutsu Toothpaste	0.012	0.1612	2.2423	0.0066
Doi Kham Carrot	Shieldtox Mosquito Spray	0.0101	0.1473	2.2312	0.0056
Shieldtox Mosquito Spray	Doi Kham Carrot	0.0101	0.1526	2.2312	0.0056

Show Top Results: 20

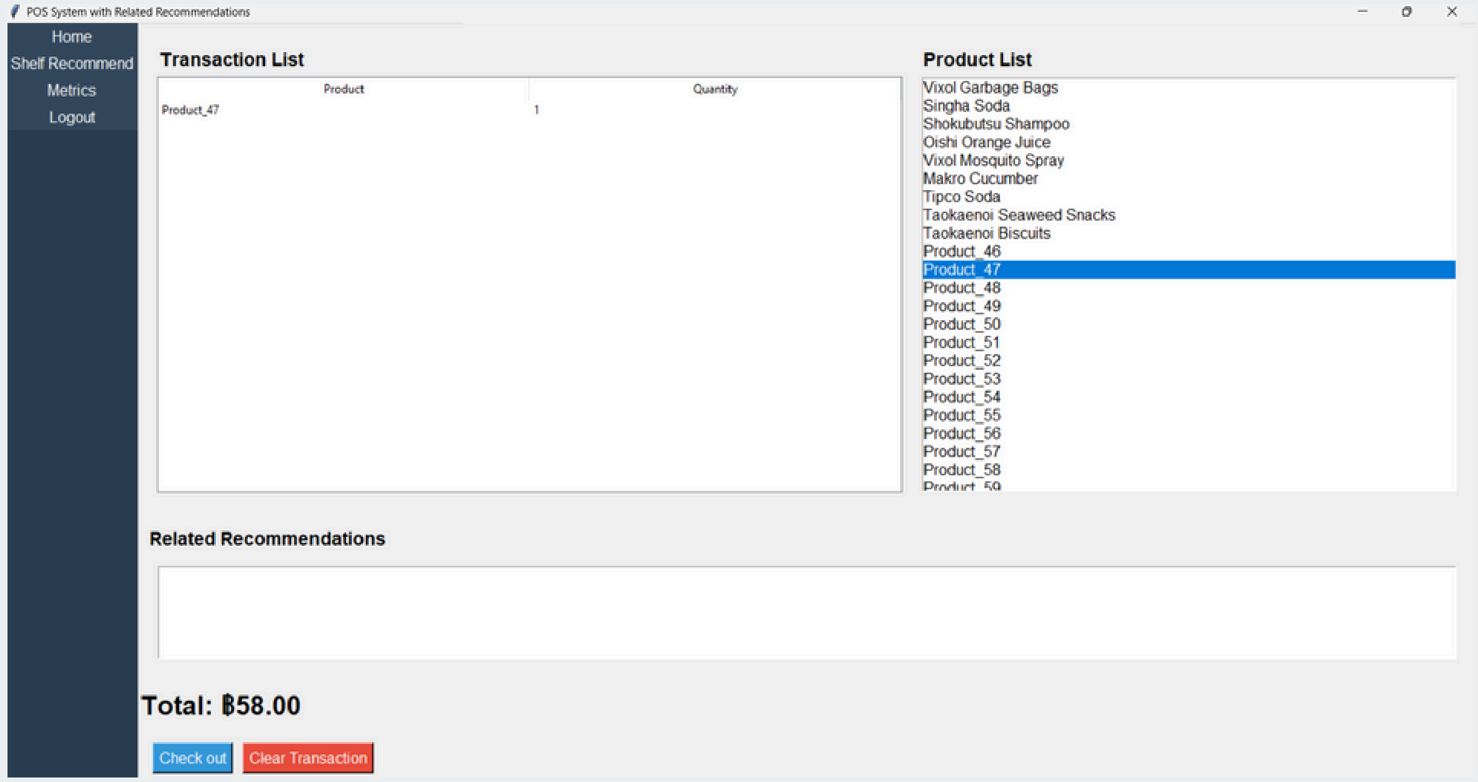
Metric Explanations

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

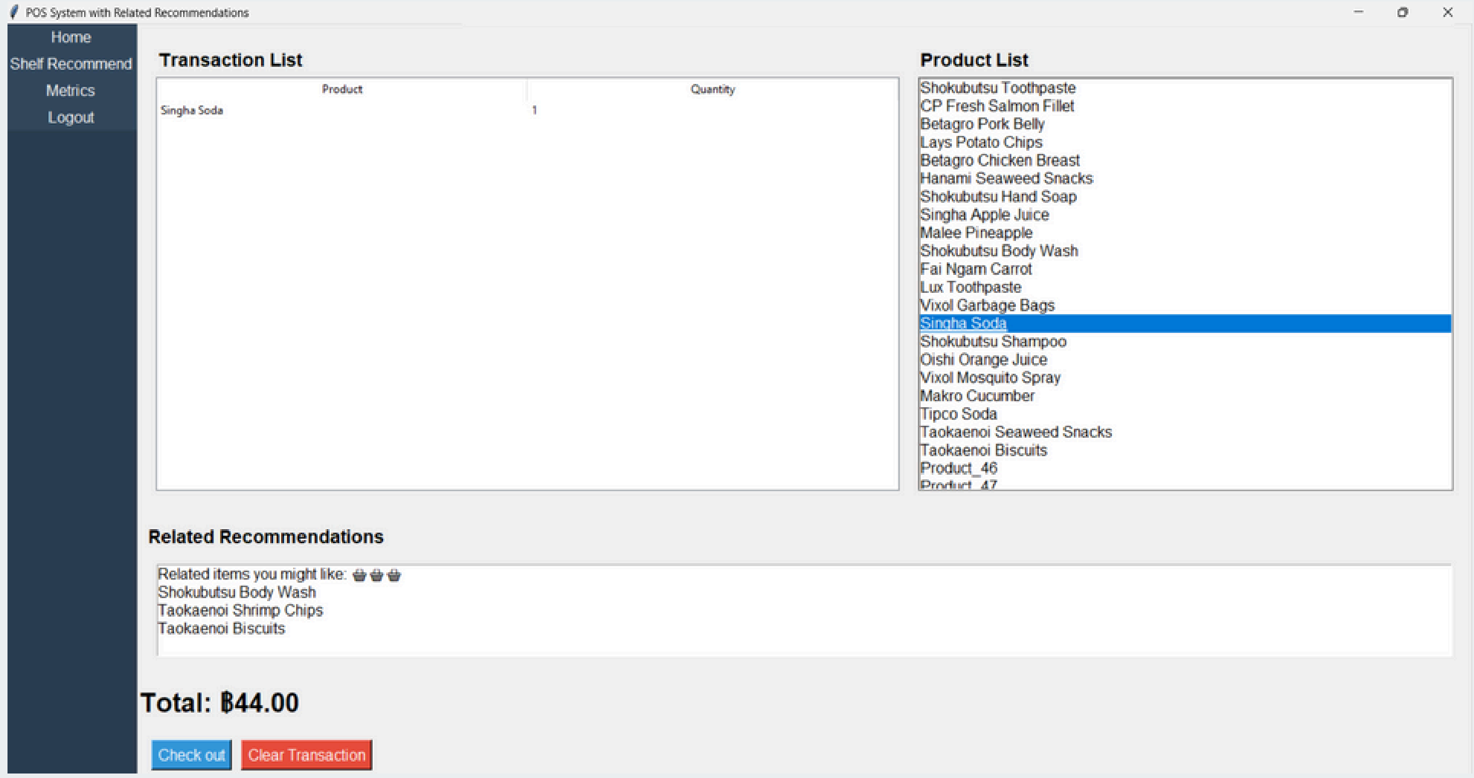
RECOMMENDATION

Real time

FUNCTION get_related_recommendations(scanned_items):
    LOAD association rules from database
    FIND rules where result match scanned items
    GENERATE recommendations based on these rules
    RETURN top 5 recommendations to human for approval



Not found

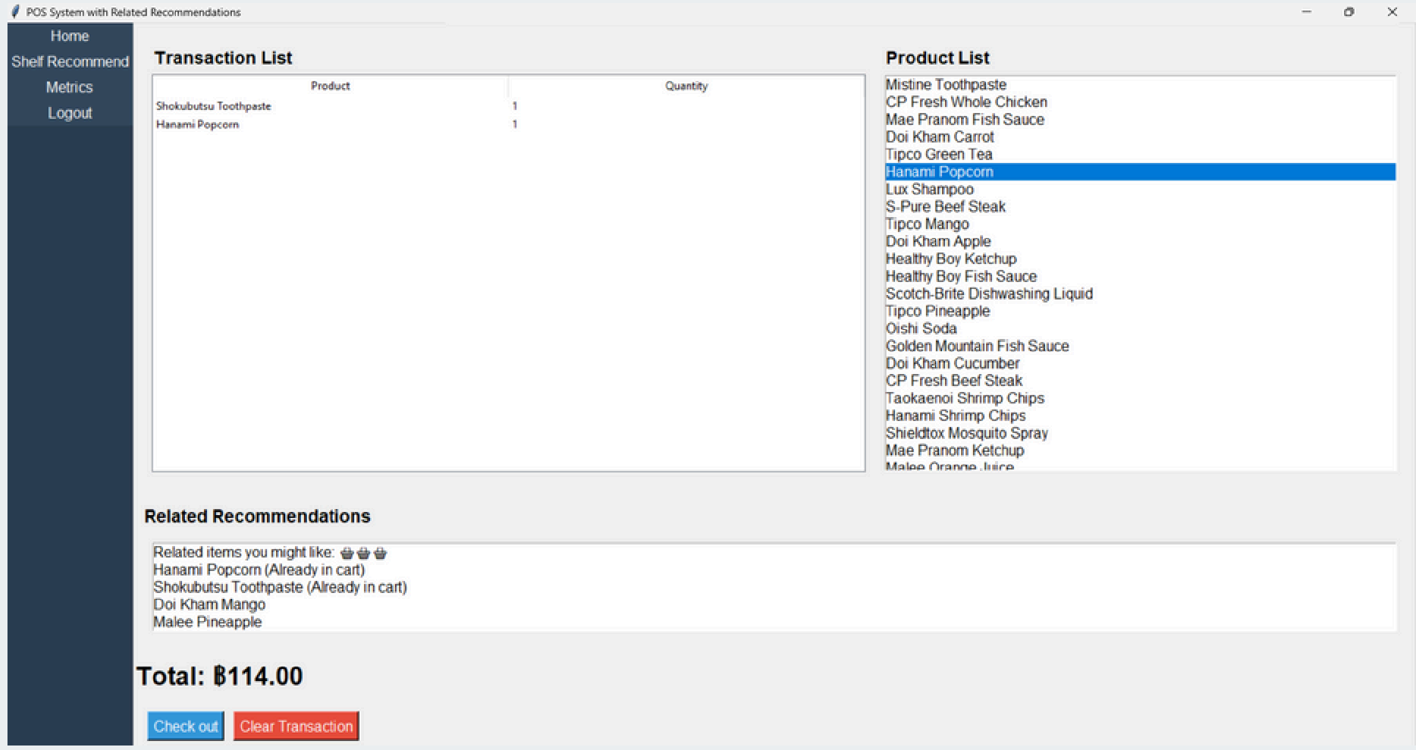


Found

RECOMMENDATION

Real time

```
FUNCTION save_log(transaction_id, recommended_items, purchased_items):
    CREATE timestamp
    INSERT transaction data (transaction_id, recommended_items, purchased_items, timestamp)
    INTO recommendation_logs
    INSERT anonymization status INTO anonymization_logs
    COMMIT logs to database
```



transaction_id	recommended_items	purchased_items	timestamp
REC0001	Hanami Popcorn, Shokubutsu Toothpaste, Doi Kham Mango, Malee	Shokubutsu Toothpaste, Hanami Popcorn	2024-10-11 09:23:24
REC0002	Scotch-Brite Dishwashing Liquid, Healthy Boy Ketchup, Doi Kham M	Hanami Shrimp Chips, Doi Kham Apple, Healthy Boy Ketchup, Scotc	2024-10-11 09:23:50
REC0003	Doi Kham Mango, Fai Ngam Carrot, Shokubutsu Body Wash, Malee (	Tipco Pineapple, Doi Kham Mango	2024-10-11 09:24:00

MRETRICS

Privacy

Percentage of customer data that is anonymized before being processed

$$\text{Anonymized Percentage} = \frac{\text{Successful Anonymization}}{\text{Total Records}} \times 100$$

Tranprarency

Percentage of recommendations with metrics provided

$$\text{Transparency Percentage} = \frac{\text{Valid Metrics}}{\text{Total Model Results}} \times 100$$

MRETRICS

Fairness

The percantage of coverage of recommendation among the customers

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

Conversion Rate (Accuracy)

Precision@K

The precision at a specific rank K, measuring the proportion of relevant recommended items in the top K recommendations

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

MRETRICS

Conversion Rate (Accuracy)

Recall@K

The recall at rank K, measuring the proportion of purchased items that are included in the top K recommended items

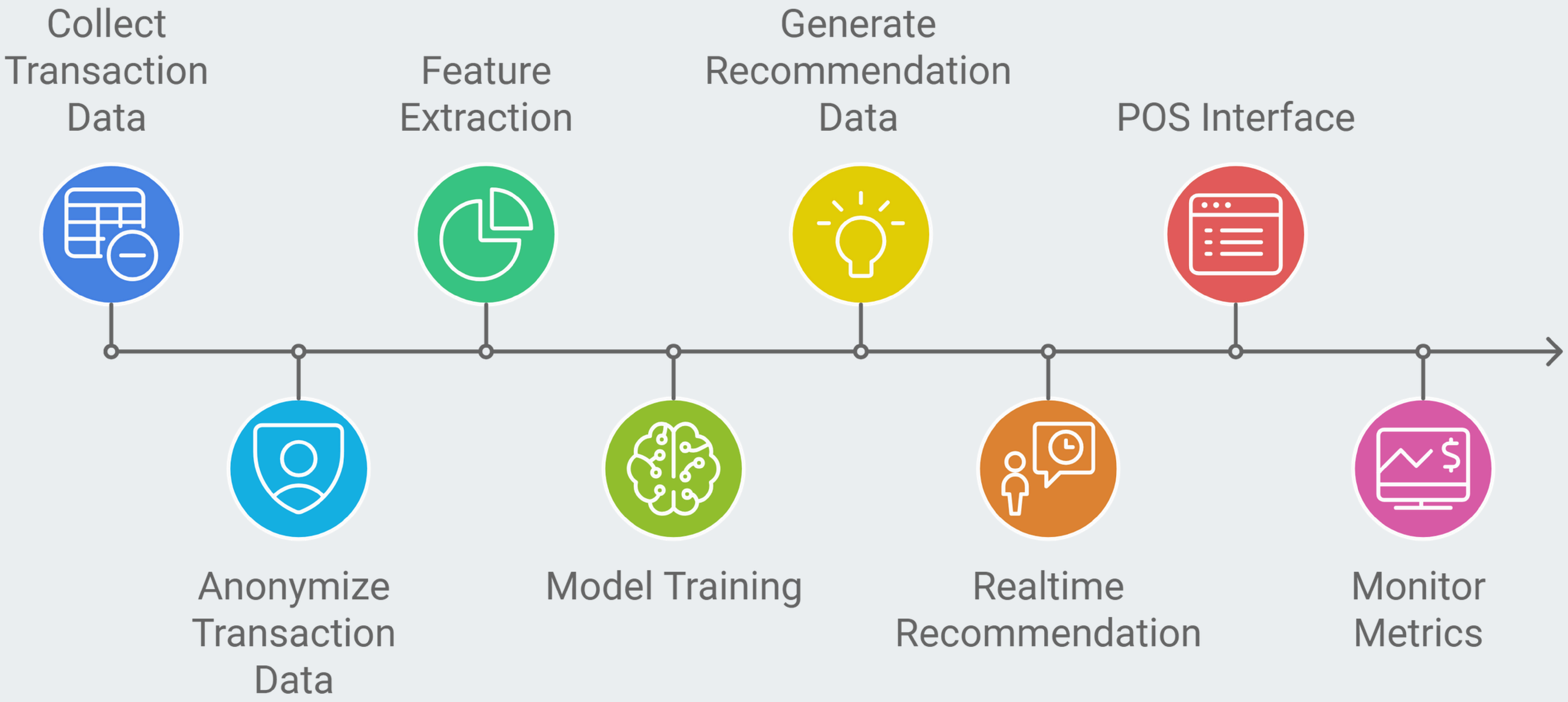
$$\text{Recall@K} = \frac{|\text{Recommended Items} \cap \text{Purchased Items}|}{|\text{Purchased Items}|}$$

Percentage Meeting Precision Threshold (PMPT):

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

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

DATA PIPELINE



APPLICATION

CONCLUSION

AI-enabled recommendation system developed for local convenience stores in Thailand not only optimizes shelf arrangements and enhances customer experiences but also fully complies with the country's AI policy framework by show how to develop the system and evaluate by several metrics to ensure that   this system is poised to boost sales while respecting customer data privacy and local buying behaviors

THANK YOU