



សាកលវិទ្យាល័យនំរតុន
NORTON UNIVERSITY

Smart E-Commerce System

with AI Recommendation Engine
& Customer Behavior Analytics

Machine Learning Course for Mid term

Subject: Machine Learning Course
| Prof. Sek Socheat





សាកលវិទ្យាល័យនំរតុន
NORTON UNIVERSITY

About our Members Group 8



Yous SETHA
Team Members



Ngeth Sereyboth
Team Members



Ouk Chanmeanrith
Team Members



Yav TOLA
Team Members

Background & Problem Statement

Limited Personalized Recommendations - Many e-commerce platforms recommend products based on popular items instead of individual customer preferences. This reduces customer satisfaction and product discovery.

Inefficient Use of Customer Data - Customer activities such as browsing and purchasing generate valuable data, but much of it is not analyzed effectively to understand customer behavior.

Difficulty Understanding Customer Behavior - Without customer behavior analytics, businesses find it difficult to identify shopping patterns, customer segments, and buying trends for better decision-making.

Need for AI-Based Solutions - Artificial Intelligence can improve product recommendations and customer behavior analysis, helping businesses provide personalized experiences and increase sales.

Project Objectives

General and specific goals of the ML system

General Objective: Build an ML-powered recommendation engine and customer behavior analytics platform that improves product discovery and provides actionable insights for Cambodian e-commerce merchants.

1

To develop an AI-powered recommendation engine for personalized product suggestions.

2

To analyze customer browsing and purchasing behavior using machine learning.

3

To identify customer segments based on shopping behavior.

4

To improve recommendation accuracy through AI techniques.

5

To provide customer behavior insights for business decision-making.

6

To evaluate the effectiveness of the AI recommendation and analytics system.

Expected Outcomes

1

An AI recommendation engine that delivers personalized product recommendations.

2

A customer behavior analytics module that identifies shopping patterns.

3

Customer segmentation based on purchasing behavior.

4

More accurate and relevant product recommendations.

5

Useful business insights through customer analytics.

6

Improved customer satisfaction and business performance.

Target Users & Beneficiaries

Who benefits from the AI recommendation system?

Five key stakeholder groups are served by this platform across the e-commerce value chain.



Online Shoppers

1

Receive personalized product recommendations and an improved, time-saving shopping experience



E-Commerce Businesses

2

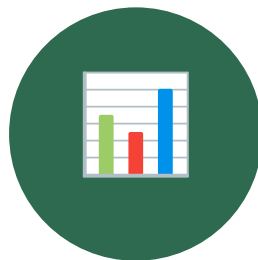
Gain actionable insights to increase conversion rates, average order value, and customer retention



Marketing Teams

3

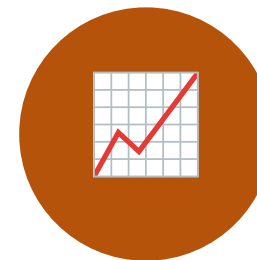
Leverage customer segments and behavior data to design and run highly targeted campaigns



Sales Managers

4

Understand product performance, demand trends, and customer preferences to guide sales strategy



Business Analysts

5

Access rich analytics reports and ML-derived insights for confident, data-driven decision making

Dataset Source & Data Collection Plan

Dataset Options

★ **Option 1 — Recommended**

UCI Online Retail Dataset

Customer ID	Invoice No.	Product ID
Quantity	Unit Price	Purchase Date
Country		

Option 2 — Kaggle Datasets

- E-Commerce Customer Behavior Dataset
- Online Retail II Dataset
- Customer Shopping Dataset
- Amazon Product Recommendation Dataset

Data Collection Plan

- 1 Source Dataset**
Download from UCI Repository or Kaggle; verify licence and format
- 2 Load into Pandas**
Read CSV files with `pd.read_csv()`; inspect shape, dtypes, and nulls
- 3 Document Structure**
Record all column names, data types, and value ranges in a data dictionary
- 4 Format for ML**
Encode categoricals, normalise numerics, handle missing values

Features & Target Variable

Input features and prediction targets for each ML model

Input Features

Customer Demographics		
Customer ID	Age	Gender
Country		
Purchase Behavior		
Purchase History	Product Category	Product Price
Quantity Purchased	Purchase Frequency	Total Spending
Time Since Last Purchase		
Product Interaction		
Product Rating		

* Feature engineering: interaction scores, recency-frequency-monetary (RFM) indices, and session-level aggregates will be derived from raw fields.

Target Variables

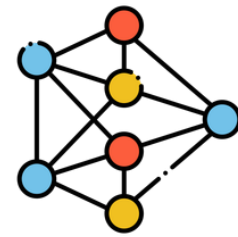
Recommendation Engine	
Next Product to Purchase	Ranking / Multi-class
Predict which product(s) the customer is most likely to buy next, ranked by probability	
Customer Analytics	
Will Purchase? (Yes / No)	Binary Classification
Binary classification — predict whether a customer will make a purchase in the next session	
Customer Segmentation	
Customer Segment	Clustering (unsupervised)
Assign each customer to a behavioural cluster (e.g. High-Value, At-Risk, New) using unsupervised ML	
3 Model Types • 12 Input Features • 3 Target Variables	

Proposed ML Algorithms



Recommendation System

- Collaborative Filtering
- Content-Based Filtering



Machine Learning Algorithms

- Random Forest
- Decision Tree
- K-Nearest Neighbors (KNN)



Customer Segmentation

- K-Means Clustering

"Overall, Collaborative Filtering handles product recommendations, Random Forest predicts purchasing behavior, and K-Means helps us understand different customer groups. Random Forest is the best choice because it is accurate, easy to explain, and widely used."

Proposed Evaluation Metrics



For Classification

- Accuracy
- Precision
- Recall
- F1 Score
- ROC-AUC



For Recommendation

- Precision@K
- Recall@K



For Clustering

- Silhouette Score

Classification metrics

Accuracy

Correct predictions / Total predictions

100 tested, 92 correct → **92%**

Precision

True positives / Predicted positives

45 of 50 predicted purchases correct → **90%**. High precision = few false positives.

Recall

True positives / Actual positives

45 of 60 actual purchasers caught → **75%**. High recall = few missed opportunities.

F1 score

$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Harmonic mean of precision and recall — useful on imbalanced data.

ROC-AUC

Ability to distinguish classes across all decision thresholds.

AUC = 1.0 — excellent model **AUC = 0.5** — random guessing

Recommendation metrics

Precision@K

Relevant items in top K / K

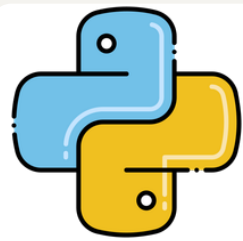
8 of top 10 recommendations relevant → **80%**

Recall@K

Relevant items in top K / Total relevant items

12 of 20 relevant products recommended → **60%**

Programming Language, Framework & Tools



Programming Language

- Python



Libraries

- Pandas
- NumPy
- Scikit-learn
- Matplotlib
- Plotly



Machine Learning

- Scikit-learn



Development Environment

- VS Code



Database

- CSV Dataset
- SQLite (optional)



Version Control

- GitHub

Expected Output of the system

AI Recommendations live

Customer Behavior Insights

**Inventory & Demand
Forecasting**

**Customer Segmentation
Engine**

**Fast reliable Platform
Responsive &
Easy to use interface**



Work Plan and timeline

- Literature review and dataset selection
- Data cleaning and preprocessing
- Feature engineering
- Model development



week
1-3

- Model training
- Model evaluation
- Recommendation engine integration



week
3-5



Week
5-6

- Dashboard development
- Testing and improvements
- Documentation and presentation

Limitations & Risks

1

Limited Real-World Data

Production interaction data not yet available; synthetic dataset will be used for development. Results may not fully reflect real behavior.

2

Cold-Start Problem

New users and new products have no interaction history, limiting collaborative filtering accuracy at launch.

3

Data Privacy

Customer behavior tracking involves sensitive data; anonymization and ethical data handling are required.

4

Model Drift

Customer preferences change over time; the model will need periodic retraining to maintain accuracy in production.



សាកលវិទ្យាល័យនំរតុន
NORTON UNIVERSITY

QUESTIONS

