Thesis Defense

— SUBMITTED BY
NGUYEN THUY KHANH

MAJOR SECTIONS OF THIS THESIS DEFENSE

- Problem Statement
- Objectives
- Scope & Limitations
- Theoretical Framework - Literature Review
- Methodology

PERSONALIZED RFM-ANALYSIS BASED RECOMMENDER by COLLABORATIVE FILTERING

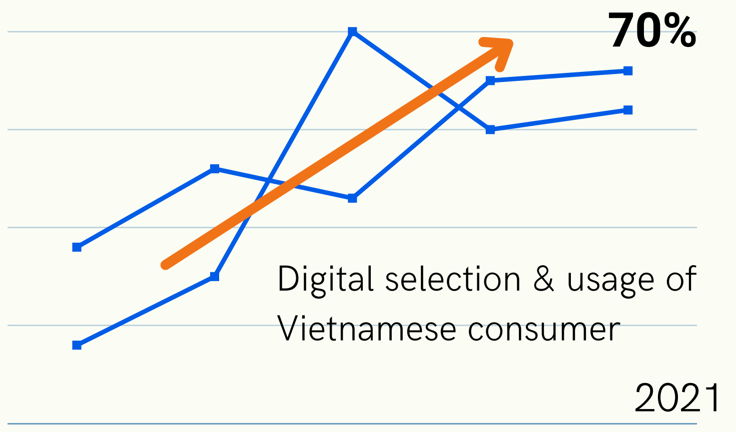
Case Study of Service Provider Y for gym & fitness

SEM I/ 2022

HCMC INTERNATIONAL UNIVERSITY

BACKGROUND

CUSTOMER BEHAVIOR TRANSFORMATION LEADS TO THE URGE OF BIG DATA ANALYTICS ADOPTION



Customer interaction: Company Y adopted an online business upgrade
- Marketing (Mailing / Mobile App): update promo/info to all users weekly to monthly
- Nearly no record improvements in user interactions, rates of active users and newcomers

Customer segmentation based on standard Marketing rules
- No customized contents delivered to particular customers

The need for an effective way to:

- target specific customer groups
- communicate with potential customers, reactivate at-risk clients and retain valued ones

The thesis promotes RS integration with other systems, in the case of firm Y, the market segmentation base is therefore proposed and implemented.

→ **Personalized Recommender system is hereby the answer**

— PART 1

Problem Statement — CASE STUDY

Objectives

THIS STUDY PROPOSES AN INTEGRATED FRAMEWORK OF A PERSONALIZED PRODUCT RECOMMENDER BASED ON WEIGHTED RECENCY, FREQUENCY, MONETARY (RFM) ANALYSIS.

- Derived priorities for each variable by the AHP method, the **weighted RFM analysis** is employed where conducts and compares **k-means** and **fuzzy c-means clustering** algorithms; then obtains the optimal number of customer segments.
- Finally, the model implements **User-based collaborative filtering** to provide product recommendations to each segment. To highlight our model benmarch, compare our model vs both the traditional CF system and Non-weighted RFM.

SCOPE

- 30,000 historical customer transactions were collected from 28 August 2020 to 19 May 2021 at all 6 Ho Chi Minh city stores
- The key assumption of our research: customers in the same segment would likely share the same purchasing behavior

LIMITATION

- Product attributes are overlooked
- Limited data to current demographical users (around the South Vietnam area)
- None customer communication trackings (via mobile application/email) except transactions -> only can measure indirectly the Real-time model performance.

Theoretical Framework — PART 2

RECOMMENDER SYSTEM (RS)

Information Filtering that allows predicting products, services, and content that users may be interested in based on information collected

RS includes Content-based (CBF) and **Collaborative Filtering (CF)**

- Two types of CF models have been researched: **memory-based** CF (**User-user /Item-item CF**) and **model-based** (e.g. **Association Rules**)

THE RENOWNED RFM BY CLUSTERING TECHNIQUES

(i) Recency (R) measures the **time** since the **last transaction**

(ii) Frequency (F) measures **number of purchases** during a specific period

(iii) Monetary (M) measures the amount of **money** that customer **spends during a period.**

WELL-KNOWN APPLICATION FOR CUSTOMER SEGMENTATION (CS)

- 3 variables describes behavior are objective
- easy to formulate and understand
- Pareto Principle is at the core base

The streams of the Literature:

- **DATA-DRIVEN MARKET SEGMENTATION**

  RFM analysis by many mining techniques for specific goals of the business fields

  Some applied weighting methods to improve the reliability of the system

- **REVISED COLLABORATIVE FILTERING**

  Mostly put efforts to ease the storage issue, sparsity of traditional CF

  Compare Approaches of CF

The literature	Approach	
	RFM analysis by Clustering techniques	Recommender system
Vohra et al. (2020). Using self organizing maps and K means clustering based on RFM model for customer segmentation in the online retail business	Combine K-Means (KM) and Self-organizing maps (SOM)	None
Imani et al. (2021). Customer Segmentation to Identify Key Customers Based on RFM Model by Using Data Mining Techniques	Weighted RFM by AHP Compare K-Means (KM) and Two steps' clustering	None
Boström (2017). Comparison of user-based and item-based collaborative filtering recommendation services.	None	User_ kNN and Item_kNN algorithms
Shi et al. (2015). An Intelligent Recommendation System based on Customer Segmentation.	FCM Clustering	Only proposed association rules approach suggestion
Rodrigues et al. (2016). Product recommendation based on shared customer's behaviour.	K-Means (KM) clustering	An association rule data mining-based recommendation
Our proposed study: The Personalized RFM analysis based recommender of CF	Weighted RFM variables by AHP Compare KM and FCM algorithms	User_ kNN algorithms with rating score integrated by RFM weighted values

Literary Review

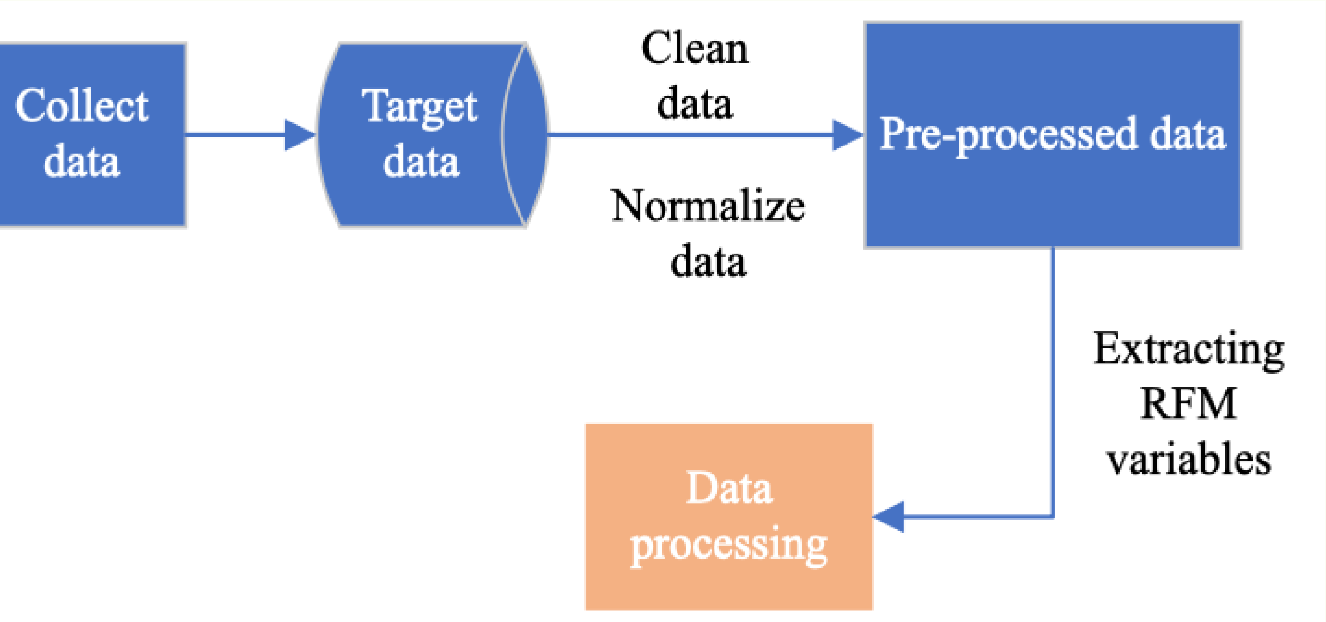
CASE STUDY PREFERENCE:

Focus: User attributes
Personalized RS suitable with local-size business resources

Methodology

PART 3

CLEAN & PROCESSING DATA



Customer number	Recency (R)	Frequency (F)	Monetary (M)
001	18	3	1900317
002	55	6	2897889

RESULT ANALYSIS

Evaluate our model

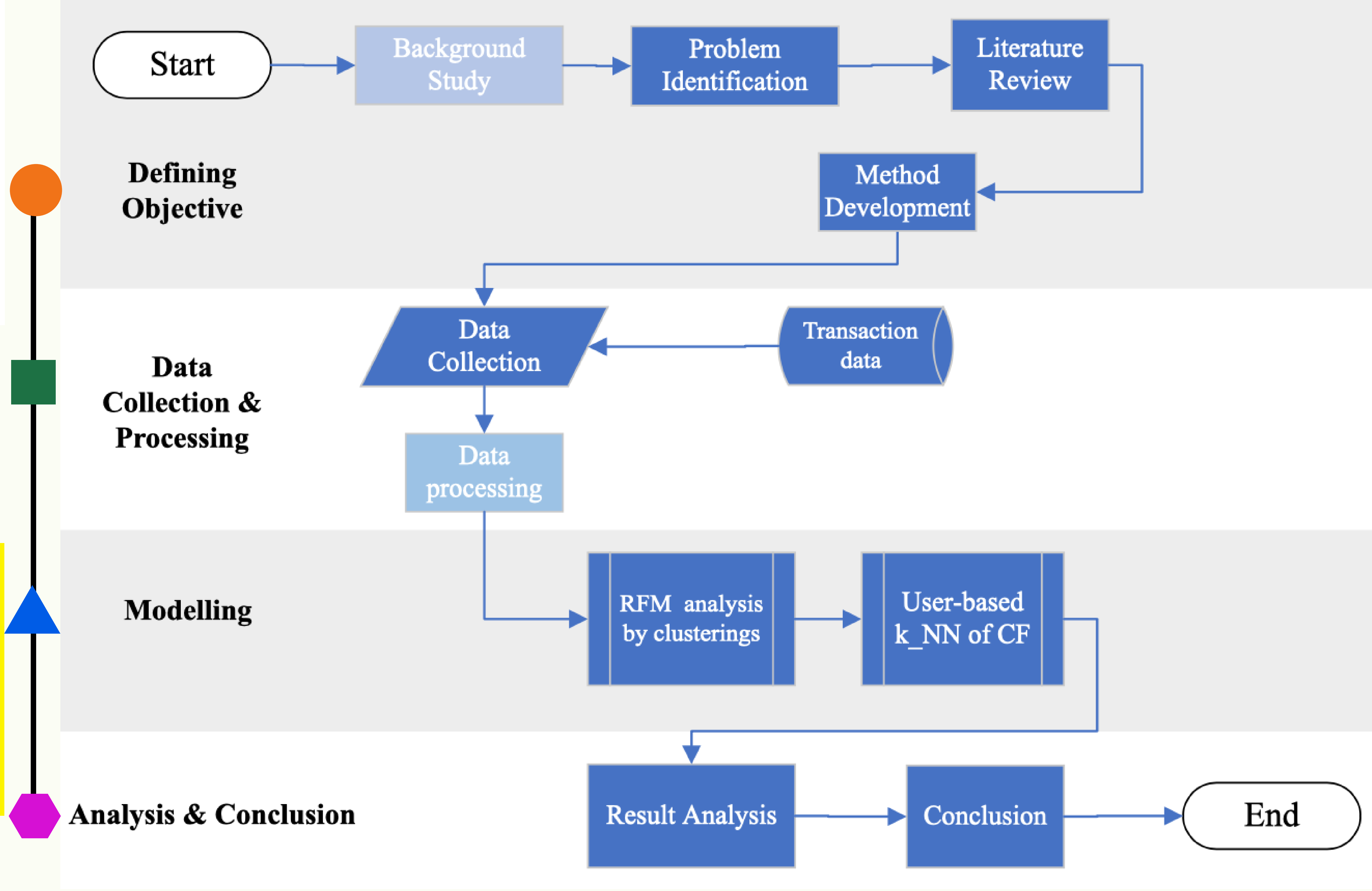
↓

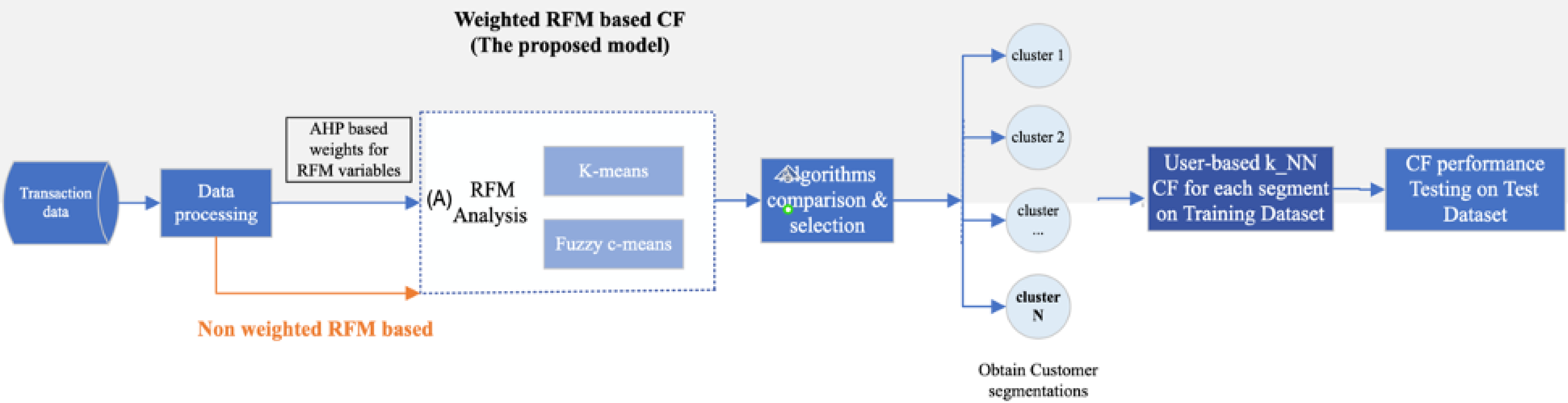
Compare & Evaluate

F1-metrics for the final model

Our model performance vs the **traditional CF model**

Our model performance vs the **non-weighted RFM model**





▲ Modeling

(A) SEGMENTATION BASED ON RFM ANALYSIS

Once obtained the number of clusters, the Silhouette index validates and compares clustering performance to select the greater alternative for this research.

(B) USER-BASED COLLABORATIVE RECOMMENDER FOR EACH SEGMENT

Apply User-kNN algorithms for computing user similarity
-> Predictions -> top N-recommendation list

TEST PERFORMANCE ON TRAINING DATA

STAGE A

Segmentation based on RFM analysis

RFM Analysis

WEIGHTED RFM VALUES EXTRACTION

calculate 3 variables weights of each customer by the AHP

Construct Pairwise Comparisons + Normalize values according to the opinion of decision makers.

Example of RFM pairwise comparison matrix

	Recency	Frequency	Monetary
Recency	1	5	7
Frequency	1/5	1	3
Monetary	1/7	1/3	1

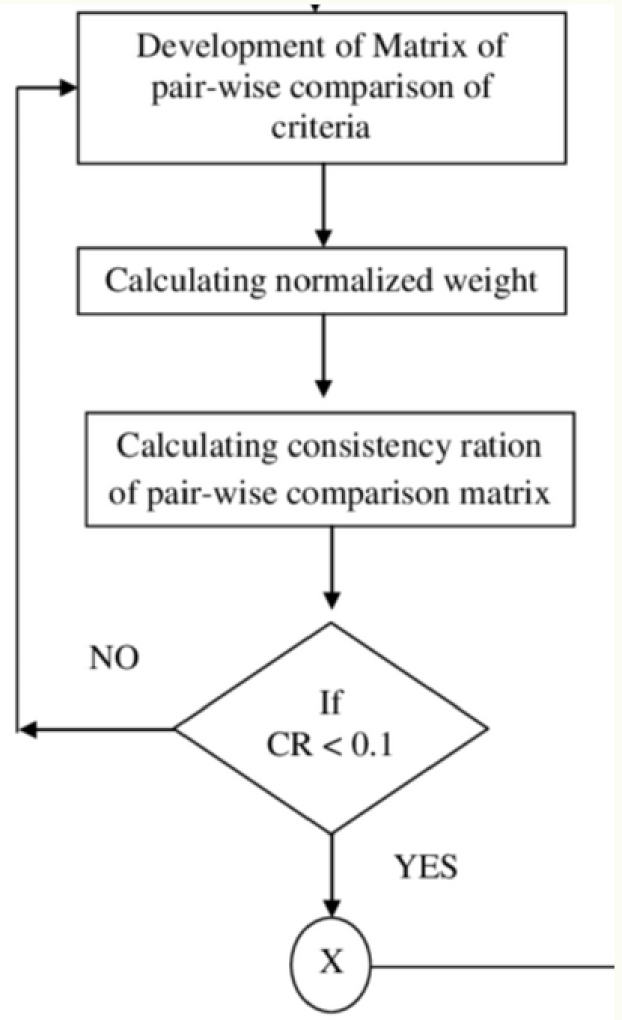
Compute the relative weights Wr - Wf - Wm

$$w_i = \frac{\sum_{j=1}^{n} b_{ij}}{n}.$$

ly giá tr thc t, c vote
t các manager. b_ij là count

Check the consistency < 0.1

Compute the consistency ratio: CR = C I / R I



Obtain weighted RFM values

Customer_id	Weighted R	Weighted F	Weighted M
000001	R x W_R	F x W_F	M x W_M
.	.	.	.
.	.	.	.
.	.	.	.

R, F, M -> RFM ca riêng user

CLUSTERINGS: KM- FCM

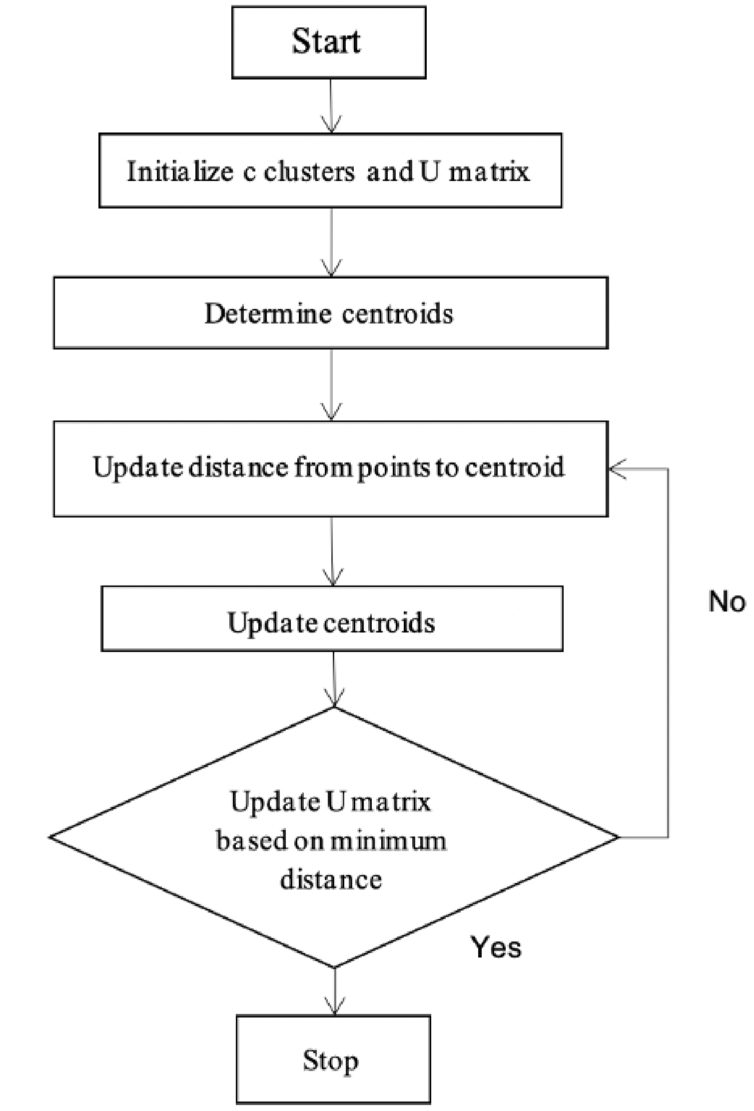
Tuning Initialize Number of clusters

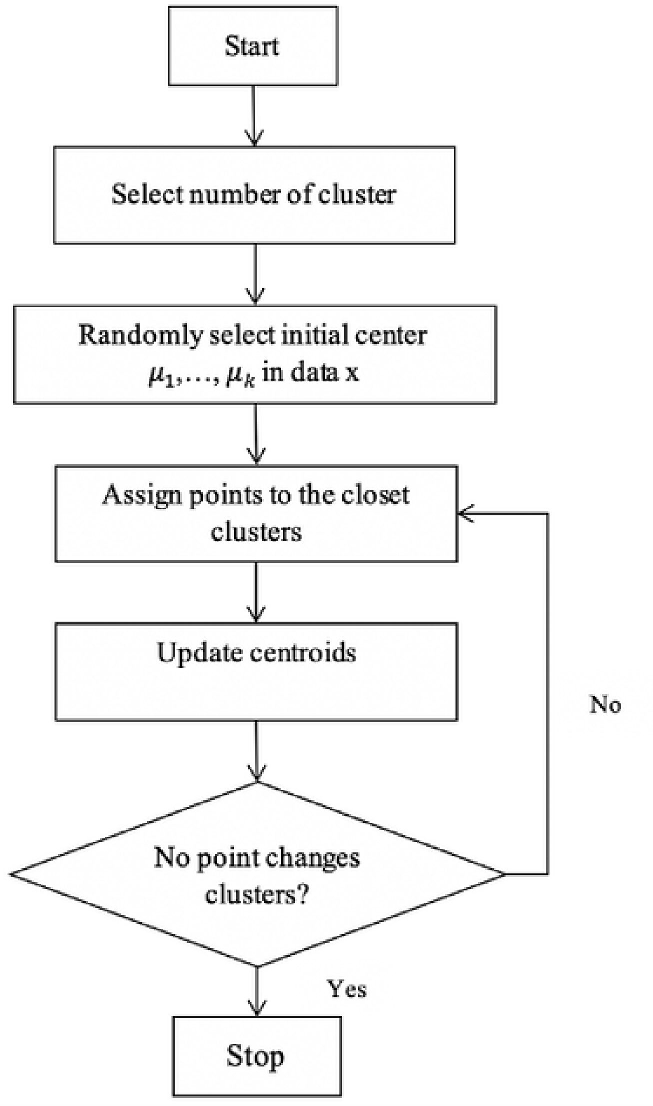
Compare & Select: K means vs **Fuzzy C-means**

Average of the Silhouette index amount for deciding about choosing optimal clusters and comparing between 2 clusterings

$$S_i = (b_i - a_i) / \max(a_i, b_i) \quad \text{for } a_i > b_i$$

The flowchart & formulas of Clustering algorithms





$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \, (4)$$

$$u_{ij} = 1 / \sum_{k=1}^{C} \left(\frac{\|x_j - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}} \, (5)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m} \, (6$$

$$\text{J(V)} = \sum_{j=1}^{k} \sum_{i=1}^{m} a_{ij} (\|x_i - v_j\|)^2 \, (1)$$

$$\mu_k = \frac{1}{n} (x^{(k_1)} + x^{(k_2)} + \cdots + x^{(k_n)}) \, (2)$$

$$c^{(i)} = argmin_k \|x_i - \mu_k\|^2 \, (3)$$

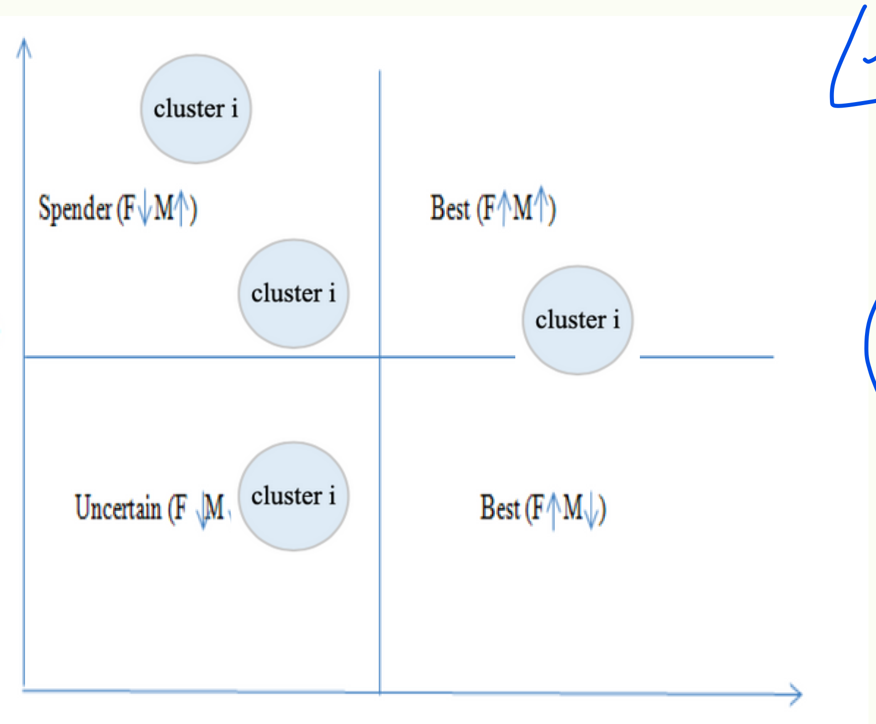
→ The core function is to minimize the sum of errors

STAGE B

Collaborative filtering each segments and recommend list

User-based CF

RFM ANALYSIS FOR SEGMENT



Apply AHP to generate weights for **each segment**

Obtain Score of each user to each item

$$C_I^j = W_R \times C_R^j + W_F \times C_F^j + W_M \times C_M^j$$

1 COMPUTING SIMILARITY

Pearson Correlation coefficient

$$sim(u,v) = \frac{\sum_{i \in I_{uv}} [(R_{ui} . \overline{R_u}).(R_{vi} - \overline{R_v})]}{\sqrt{\sum_{i \in I_{uv}} (R_{ui} - \overline{R_u})^2} \sqrt{\sum_{i \in I_{uv}} (R_{vi} - \overline{R_v})^2}}$$

0 THE USER-ITEM MATRIX

	$Item_1$	⋯	$Item_j$	⋯	$Item_n$
U_1	$R_{1,1}$	⋯	0	⋯	$R_{1,n}$
⋮	⋮	⋮	⋮	⋮	⋮
U_i	$R_{i,1}$	⋯	$R_{i,j}$	⋯	0
⋮	⋮	⋮	⋮	⋮	⋮
U_m	$R_{m,1}$	⋯	0	⋯	$R_{m,n}$

2 COMPUTING PREDICTIONS

$$P_{u,i} = \overline{R_u} + \frac{\sum_{v \in N} sim(u,v)(R_{vi} - \overline{R_v})}{\sum_{v \in N} sim(u,v)}$$

R_ij -> RFM ca user-item

3 TOP-N RECOMMENDATION SET WITH THE HIGHEST PREDICTED SCORE

STAGE A -> RFM ANALYSIS: CLUSTERS PATTERNS

Each cluster represents a market-segmentation.

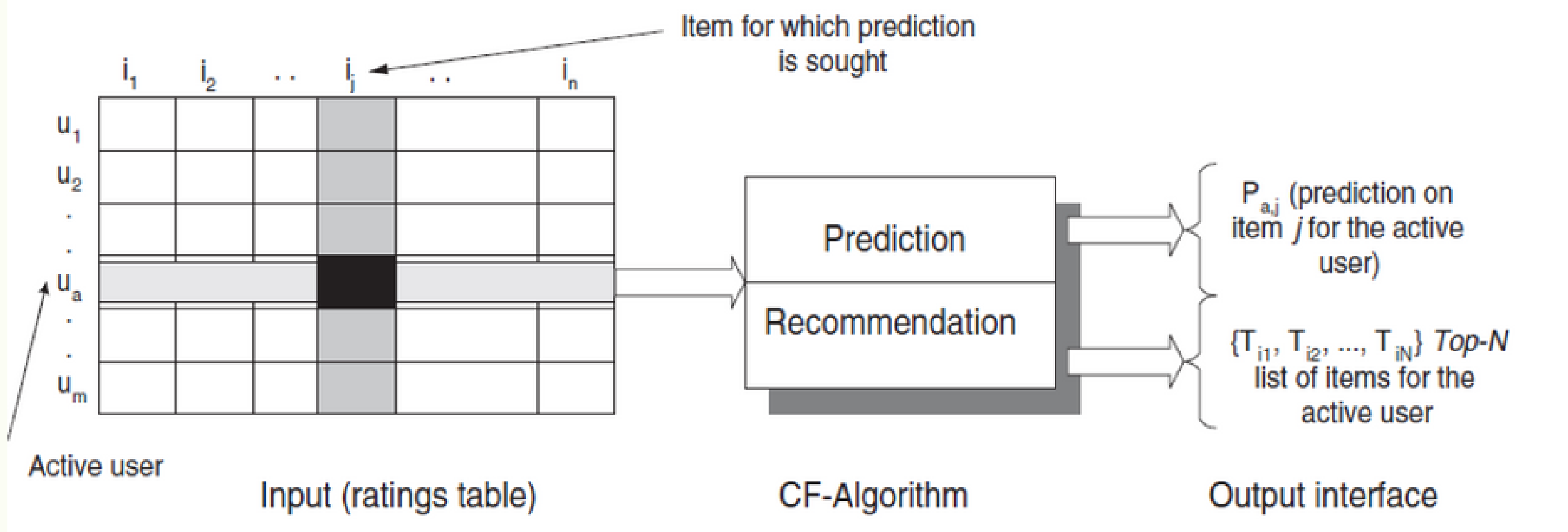
RFM analysis + AHP weights for each segment:

↓

STAGE B: RECOMMENDATION

↓

THE K-NEAREST NEIGHBOR MODEL (USER_KNN)



↓

PERFORMANCE MEASUREMENT: F1-METRICS

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Q&A

PERSONALIZED RFM-ANALYSIS BASED RECOMMENDER by COLLABORATIVE FILTERING

- RECOMMENDER SYSTEM BASED ON CUSTOMER SEGMENTATION
- RFM ANALYSIS
- CLUSTERING ALGORITHMS
- USER-BASED COLLABORATIVE FILTERING

Sincerely Thank for your listening !

ACTIVITY	START	FINISH	DURATION	WEEKS																			
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Choosing topic	1	1	0																				
Chapter I: Introduction																							
Background	1	1	0																				
Problem statement	2	2	0																				
Objectives	2	3	1																				
Scope and Limitations	2	3	1																				
Chapter II: Related works																							
Overview	4	4	0																				
Literature review	4	7	3																				
Design Concepts Considered	5	6	1																				
Chap III: Methodology																							
Approaches Comparison and Selection	7	8	1																				
Conceptual Design Description	8	8	0																				
Chapter IV: Modelling																							
Model or Prototype Development	10	15	5																				
Solution Development	13	15	2																				
Chapter V: Result Analysis																							
Results	16	18	2																				
Analysis	18	19	1																				
Chap VI: Conclusion																							
Conclusion	20	20	0																				
References	4	20	16																				

Thesis timeline