សាកលវិទ្យាល័យ ភូមិន្ទភ្នំពេញ
ROYAL UNIVERSITY OF PHNOM PENH

Recommendation System Application Development by using Association Analysis Apriori Algorithm

SAO Kimsong

Advisor: Dr. SRUN Sovila

Contents

Introduction

- E-commerce and retail companies are using the power of data and boost sales by implementing Recommendation System (RS) on their websites.
- What is a RS?
- Why do we need RS?

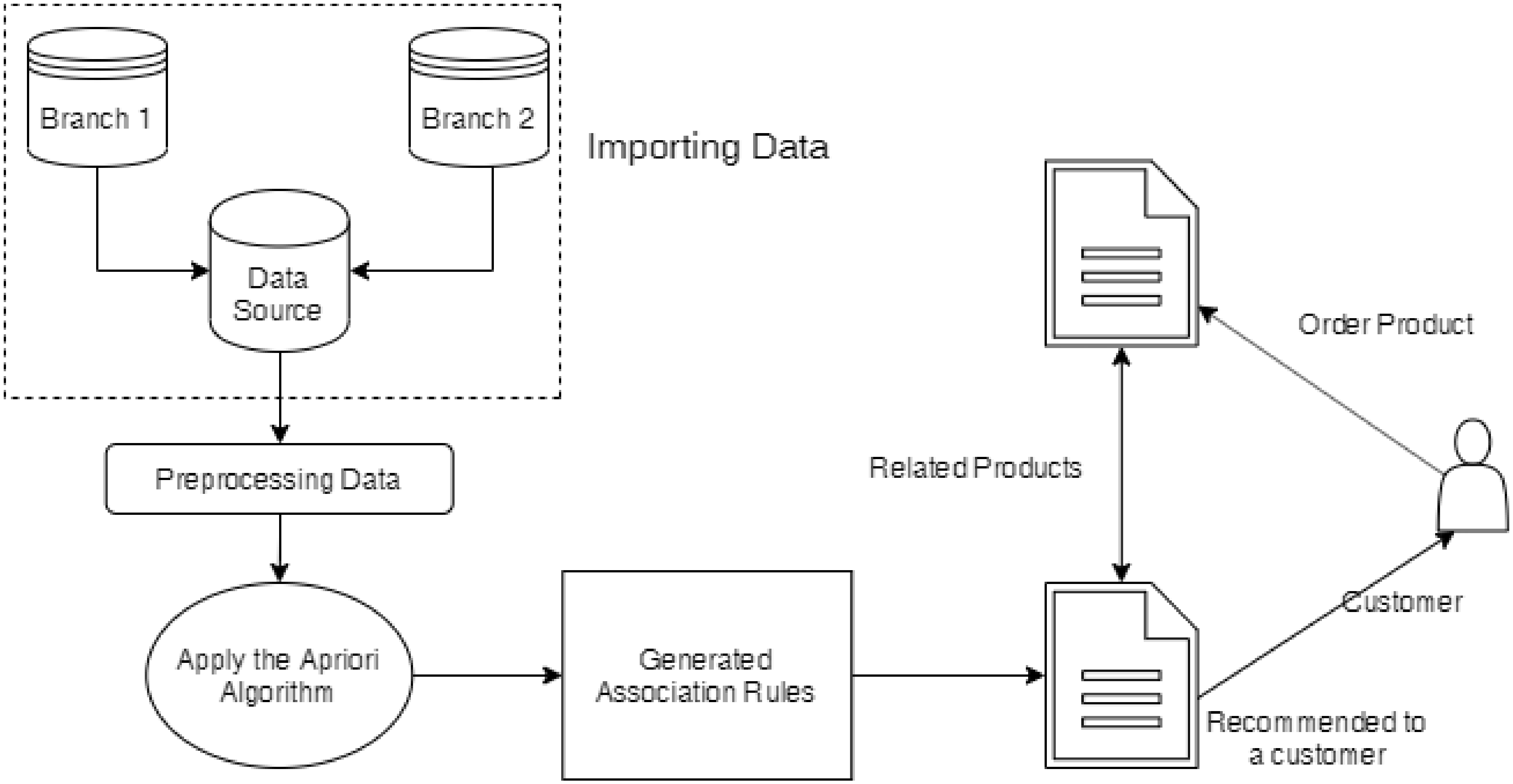
- Aims of the Study
  - Proposed the architecture of association item analysis for the RS.
  - Developed and conducted experiments of RS by using Apriori Algorithm.

Literature Reviews

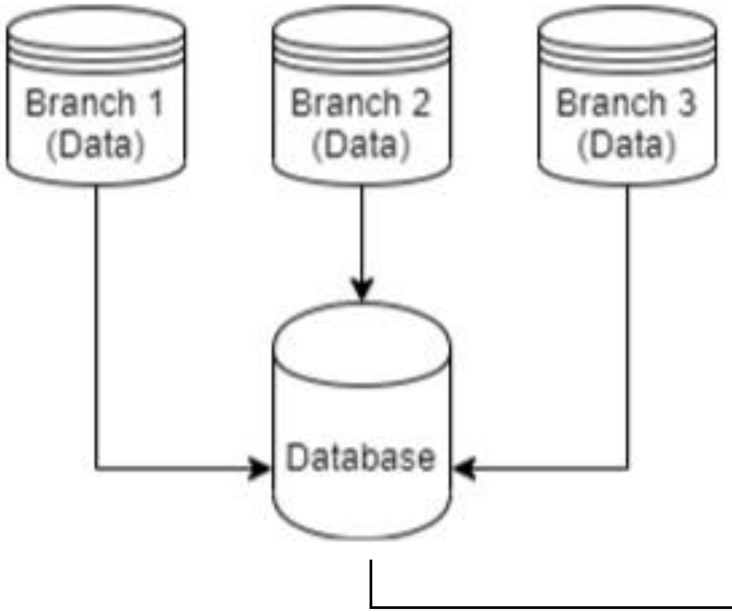
- Bendakir and Aimeur, 2006: *Proposed a course recommendation system based on association rules for students.*
- Chellatamilan, 2011: *Proposed an idea for building a recommendation system for the e-Learning system.*
- JinHyun, et al., 2016: *Implemented the mobile coupon recommendation system.*
- Shadi, et al., 2018: *Proposed a new recommender framework for requirements engineering.*
- Aijaz, et al., 2018: *Proposed technique for recommender system be using Opinion Based.*

Methodology

- System Overview



- Importing Data

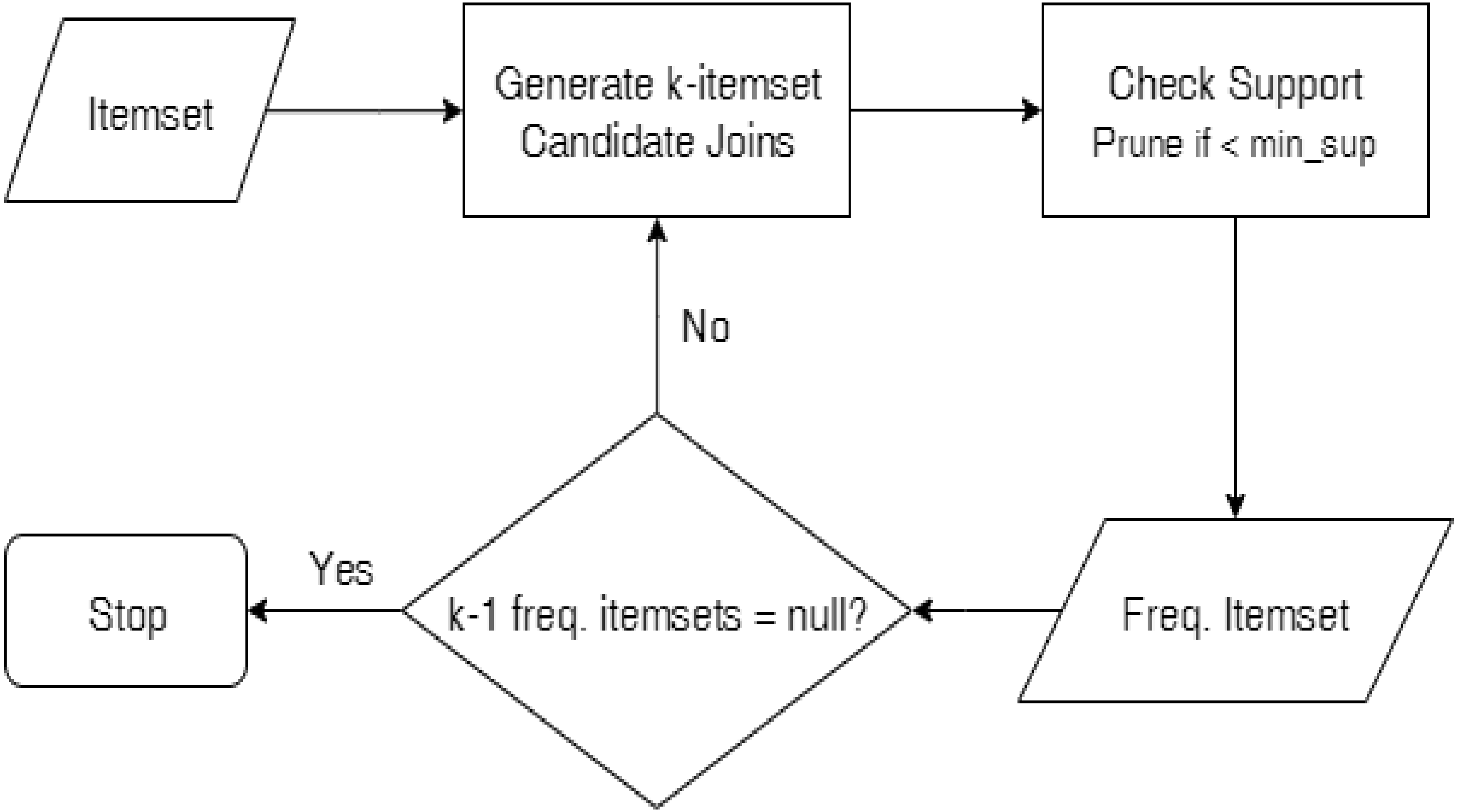


TID	Item
T1	ESPRESSO
T1	SUGAR
T1	NEWSPAPER
T2	ESPRESSO
T2	SUGAR
T2	COLA
…	…

- Preprocessing Data
  - Format transaction data to algorithm formation.
  - Labeled the item as a number.

TID	Items	Item Label
T1	ESPRESSO, SUGAR, NEWSPAPER	1, 2, 3
T2	ESPRESSO, SUGAR, COLA	1, 2, 4
T3	ESPRESSO, SUGAR	1, 2
T4	CAPPUCCINO, CIGARETTES	5, 6
T5	CAPPUCCINO, SUGAR	5, 2
T6	CAPPUCCINO, SUGAR, SWEETS	5, 2, 7
T7	DECAF, SUGAR, CHEWING_GUMS	8, 2, 9
T8	DECAF, SODA, VINEGAR	8, 10, 11
T9	DECAF, SUGAR, CIGARETTES	8, 2, 6

- Frequent Itemset for Apriori Algorithm



Methodology – Cont.

➢ Frequent Itemset for Apriori Algorithm – Example.

1-Itemsets

Item	Count	Support
ESPRESSO(1)	3	30%
SUGAR(2)	7	77%
NEWSPAPER(3)	1	11%
COLA(4)	1	11%
CAPPUCCINO(5)	3	30%
CIGARETTES(6)	2	22%
SWEETS(7)	1	11%
DECAF(8)	3	30%
CHEWING_GUMS(9)	1	11%
SODA(10)	1	11%
VINEGAR(11)	1	11%

Minimum Support = 20%

$$Support(A) = \frac{Number\ of\ transactions\ contraing\ A}{Total\ transaction}$$

2-Itemsets

Itemset	Count	Support
ESPRESSO(1), SUGAR(2)	3	30%
ESPRESSO(1), CAPPUCCINO(5)	0	0%
ESPRESSO(1), CIGARETTES(6)	0	0%
ESPRESSO(1), DECAF(8)	0	0%
SUGAR(2), CAPPUCCINO(5)	2	22%
SUGAR(2), CIGARETTES(6)	0	0%
SUGAR(2), DECAF(8)	2	22%
CAPPUCCINO(5), CIGARETTES(6)	1	11%
CAPPUCCINO(5), DECAF(8)	0	0%
DECAF(8), CIGARETTES(6)	0	0%

Itemsets removed because its support count less than minimum support count.

Methodology – Cont.

- Rule Generation

Minimum Confidence = 60%

$$Confidence(A \Longrightarrow B) = \frac{\sum transaction\ contain\ A\ \&\ B}{\sum transactions\ contain\ A}$$

Rules	Rules	Support	Confidence
$\{ESPRESSO\} \Rightarrow \{SUGAR\}$	$\{1\} \Rightarrow \{2\}$	3/9 = 30%	3/3 = 100%
$\{DECAF\} \Rightarrow \{SUGAR\}$	$\{8\} \Rightarrow \{2\}$	2/9 = 22%	2/3 = 66%
$\{CAPPUCCINO\} \Rightarrow \{SUGAR\}$	$\{5\} \Rightarrow \{2\}$	2/9 = 22%	2/3 = 66%

Experiments

➢ Experiment Setup

Hardware	Software
CPU: Intel Core i5-5200U 2.20GHz, 2 Core(s)	Windows 10 x64 Enterprise
RAM: 16GB	Python 3.7
	PyQT5, QT Designer
	Visual Studio Code

Data	Total Transaction
Super Market	4, 444
Book Store	
Super Market	16, 466

Experiments – Cont.

➢ Graphical User Interface

Freq. Itemsets

Filter Itemsets

Contains:

☑ Show Only Support >= Min. Support

Filter

Information

From Date: 09/01/2019
To Date: 10/31/2019
Min. Support (%): 0.1000
Min. Confidence (%): 10.00

	Itemsets	Transactions	Support Count	Support (%)
545	{'5514', '9702'}	5273	6	0.113787217902522290
546	{'5514', '3002'}	5273	15	0.284468044756305700
547	{'10449', '10484'}	5273	19	0.360326190024653940
548	{'10476', '10481'}	5273	6	0.113787217902522290
549	{'12658', '10341'}	5273	7	0.132751754219609330
550	{'10449', '10477'}	5273	7	0.132751754219609330
551	{'11522', '11523'}	5273	24	0.455148871610089160
552	{'13', '10272'}	5273	6	0.113787217902522290
553	{'10486', '10477'}	5273	7	0.132751754219609330
554	{'4070', '4073'}	5273	6	0.113787217902522290
555	{'12792', '3059'}	5273	11	0.208609899487957520
556	{'3059', '5514'}	5273	10	0.189645363170870470
557	{'3059', '3002'}	5273	6	0.113787217902522290
558	{'10038', '5514'}	5273	7	0.132751754219609330
559	{'10341', '10302'}	5273	20	0.379290726341740940
560	{'13165', '10272'}	5273	9	0.170680826853783440
561	{'13165', '380'}	5273	14	0.265503508439218660
562	{'10480', '10476'}	5273	25	0.474113407927176200
563	{'9609', '3204'}	5273	6	0.113787217902522290

- Results and Discussions

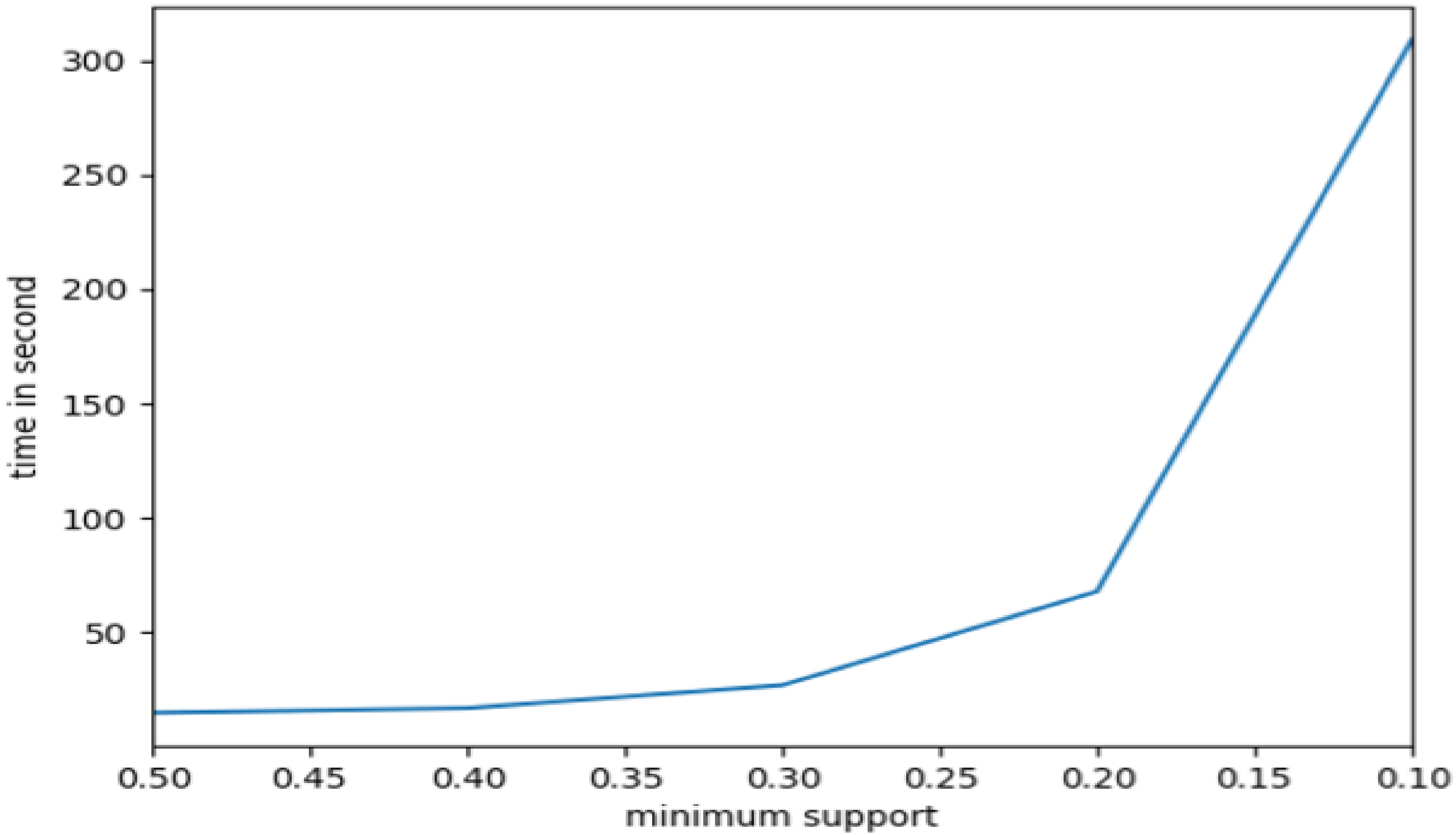


Figure 6. Response time of frequent itemset generation for Dataset2.

Conclusions & Future Works

- Conclusions
  - Proposed an architecture for association item analysis for RSs.
  - Developed and conducted experiment of RS by using Association Analysis Apriori Algorithm.
  - The results can provide recommended a new item to customers by understanding historical transaction data.
- Future Works
  - Make a library for recommend product to customers by using association items from our proposed frameworks.

Thank You

Q&A