Practicum Case	
COMP6140 | COMP6140001 | COMP6140049 Data Mining	
Computer Science	**O191-COMP6140-NP01-04**
Valid on *Even Semester Year 2019/2020*	**Revision 00**

Learning Outcome

- LO1 – explain concept of data and data preprocessing

Topic

- Session 04 – Mining Frequent Pattern and Associations Using R

Sub Topics

- Apriori
- Frequent Pattern Growth (FP-Growth)
- Association Rule
- Review

Soal
Case

Bluejack Game Store

Bluejack Game Store is a multinational game vendor which sells all of the most interesting games to play in every existing game platform. To boost it sales, the store plans to hold a promotional discount for their most popular games. It is also possible to hold discount on paired or multiple purchase of games as a package. As a data scientist, you are asked to analyze the transactions data of the company and conclude the games frequently bought together or alone to support the discount event.

You are given three documents, which are **Detail.csv** (containing detail transaction of game purchases), **Header.csv** (containing header transaction of game purchases), and **Game.csv** (containing the list of available games sold). Take note that due to some technical issues, some of the data are **invalid**. These data should be **omitted** from the analysis, as it might cause confusion to the final decision of the discount event.

GameId	Name	Platform	Year	Genre	Price	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
1	Wii Sports	Wii	2006	Sports	13.32	Nintendo	41.49	29.02	3.77	8.46	82.74
2	Super Mario Bros.	NES	1985	Platform	12.45	Nintendo	29.08	3.58	6.81	0.77	40.24
3	Mario Kart Wii	Wii	2008	Racing	15.84	Nintendo	15.85	12.88	3.79	3.31	35.82
4	Wii Sports Resort	Wii	2009	Sports	19.06	Nintendo	15.75	11.01	3.28	2.96	33
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	11.56	Nintendo	11.27	8.89	10.22	1	31.37
6	Tetris	GB	1989	Puzzle	16.25	Nintendo	23.2	2.26	4.22	0.58	30.26
7	New Super Mario Bros.	DS	2006	Platform	13.25	Nintendo	11.38	9.23	6.5	2.9	30.01
8	Wii Play	Wii	2006	Misc	11.39	Nintendo	14.03	9.2	2.93	2.85	29.02
9	New Super Mario Bros. Wii	Wii	2009	Platform	17.75	Nintendo	14.59	7.06	4.7	2.26	28.62
10	Duck Hunt	NES	1984	Shooter	17.69	Nintendo	26.93	0.63	0.28	0.47	28.31
11	Nintendogs	DS	2005	Simulation	12.87	Nintendo	9.07	11	1.93	2.75	24.76
12	Mario Kart DS	DS	2005	Racing	11.11	Nintendo	9.81	7.57	4.13	1.92	23.42
13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	14.18	Nintendo	9	6.18	7.2	0.71	23.1
14	Wii Fit	Wii	2007	Sports	11.02	Nintendo	8.94	8.03	3.6	2.15	22.72
15	Wii Fit Plus	Wii	2009	Sports	18.53	Nintendo	9.09	8.59	2.53	1.79	22

Figure 1. Game.csv Data

TransactionId	TransactionDate
TR0022	4/27/2018
TR0033	6/1/2018
TR0062	3/3/2018
TR0063	5/9/2018
TR0114	4/8/2018
TR0126	1/12/2018
TR0135	4/4/2018
TR0154	5/11/2018
TR0156	5/17/2018

Figure 2. Header.csv Data

TransactionId	GameId	Quantity
TR0022	4	4
TR0022	22	3
TR0022	34	3
TR0022	44	4
TR0022	76	4
TR0033	1	5
TR0033	3	1
TR0033	8	1
TR0033	56	2

Figure 3. Detail.csv Data

To **analyze** the data, you need to do the following steps:

1. **Apriori Analysis**
   a. **Remove invalid** transaction data.
   b. **Prepare** the **data** for Apriori.
   c. Perform **Apriori analysis** with **minimum support 0.005** to produce **frequent itemsets**.

```
      items                                  support count
[1]   {Call of Duty: World at War}       0.042381433    42
[2]   {Donkey Kong Country}              0.042381433    42
[3]   {Pac-Man}                          0.045408678    45
[4]   {Mario Kart 64}                    0.045408678    45
[5]   {Wii Party}                        0.043390515    43
[6]   {Battlefield 3}                    0.046417760    46
[7]   {Mario Party 8}                    0.043390515    43
[8]   {New Super Mario Bros. 2}          0.045408678    45
[9]   {Call of Duty: Ghosts}             0.044399596    44
[10]  {Halo 2}                           0.046417760    46
```

Figure 4. Frequent Itemsets

   d. Produce **Association Rules** from the frequent itemsets with **minimum confidence 0.15**.

```
    lhs                     rhs                   support     confidence lift
[1] {Super Mario World} => {Super Mario Kart}  0.007063572 0.1521739  3.278355
[2] {Super Mario Kart}  => {Super Mario World} 0.007063572 0.1521739  3.278355
```

Figure 6. Association Rules

2. FP Growth Analysis

a. **Remove invalid** transaction data.

b. **Prepare** the data for **FP Growth** by transforming it into the right format.

1	16	25	47	74	76
1	29	34	61	79	
1	5	26	50	75	
1	49	67			
1	4	11	25	26	
1	6	10	15	85	
1	37	42	65	66	93
1	25	78	79	82	
1	14	72	100		
1	9	40	85		
1	33	73	92		
1	79	84	86	98	

Figure 7. Example of Data Accepted by FP-Growth

c. Perform **FP Growth analysis** with **minimum support 0.15** to produce **frequent itemsets**.

```
16 (4.995)
47,16 (0.2997)
47 (4.995)
32,16 (0.3996)
32,47 (0.1998)
32 (4.995)
39,16 (0.3996)
39,47 (0.1998)
39 (4.995)
29,16 (0.1998)
29,32 (0.0999001)
29,47 (0.0999001)
29 (4.995)
30,29 (0.3996)
30,47 (0.3996)
30,32 (0.2997)
30,39 (0.0999001)
30,16 (0.0999001)
30 (4.995)
12,16 (0.4995)
12,32,16 (0.0999001)
12,32 (0.2997)
```

Figure 8. FP-Growth Frequent Itemsets Result