|  |  |
| --- | --- |
| **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.



***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:

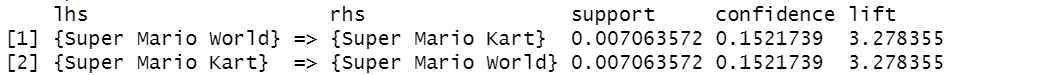
1. **Apriori Analysis**
   1. **Remove** **invalid** transaction data.
   2. **Prepare** the **data** for Apriori.
   3. Perform **Apriori analysis** with **minimum support 0.005** to produce **frequent itemsets**.

**A picture containing graphical user interface

Description automatically generated**

***Figure 4. Frequent Itemsets***

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



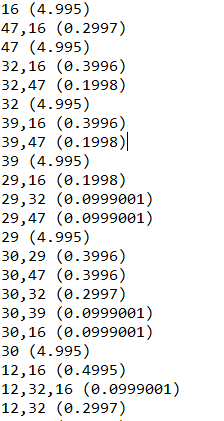
***Figure 6. Association Rules***

1. **FP Growth Analysis**
   1. **Remove** **invalid** transaction data.
   2. **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***

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



***Figure 8. FP-Growth Frequent Itemsets Result***