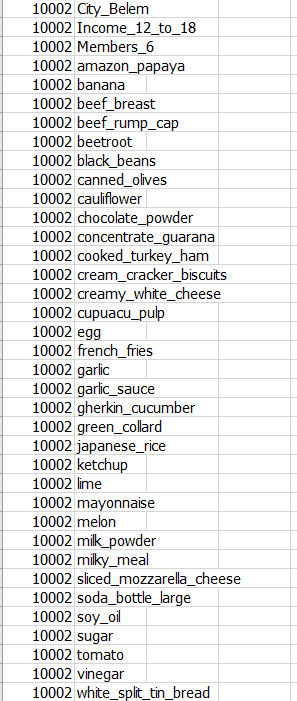
**EGCO 425 – Project 1 (Association)**

**ตอนที่1**

CES Dataset เป็นฐานข้อมูลที่ได้มาจากการสำรวจ 1540 ครอบครัวจาก 8 เขต ที่อยู่ในที่อยู่อาศัยในบราซิล ในการไปซื้อของในห้างสรรพสินค้า โดยเก็บข้อมูลเป็น เมือง,จำนวนสมาชิกในครอบครัว,เงินเดือนและรายการซื้อสินค้า โดยเรียกแบบสำรวจนี้ว่า Consumer Expenditure Survey

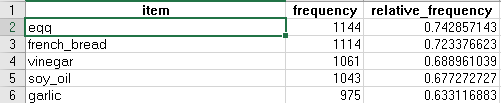
โดยจะมี Transaction id เป็นตัวแสดงถึงรายการ การซื่อของในห้างสรรพสินค้าของครอบครัวหนึ่ง โดย รายการแรกบอกถึงเมืองที่อาศัย รายการสองบอกถึงเงินเดือนที่เหลือและรายการสุดท้ายคือจำนวนสมาชิกในครอบครัว(ถ้าเป็นหนึ่งแสดงว่าอยู่คนเดียว) และรายการที่เหลือเป็นรายการในการซื้อสินค้า

**ยกตัวอย่าง**



ครอบครัวนี้อาศัยอยู่ในเมือง Belem มีรายได้12 ถึง18 ต่อเดือน มีจำนวนสมาชิกในครอบครัว 6 คน จะทำการซื้อสินค้าพวก มะละกอ กล้วย เนื้อหน้าอก ไข่ และอื่นๆ

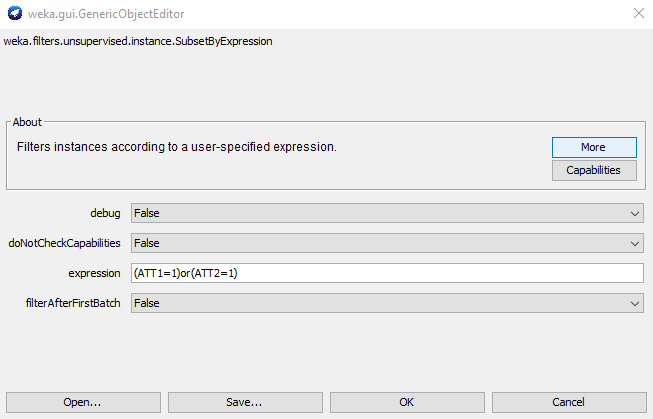
โดยตัวแปร city กับ สินค้า จะเป็น Nominal เพราะจะแยกความต่างจากตัวอักษร ตัวแปร income จะเป็นตัวแปร Ordinal เพราะมีการแยกค่าเป็นช่วงโดยที่ไม่ทราบค่าจริง เช่น Income\_12to\_18 จะรู้แค่ว่ารายได้ต่อเดือนของครอบครัวนี้เป็น 12ถึง18 แต่ไม่ทราบค่าเงินเดือนจริงๆ อาจจะเป็น 13หรือ15ก็ได้ ตัวแปร member เป็น interval เพราะความต่างของจำนวนสมาชิกครอบครัวมีผลต่อการทำ association และไม่มีทางที่จำนวนสมาชิกครอบครัวเป็น 0

****

จากค่าสถิติที่เราเห็น จะสามารถทำนายได้ว่าหลังจากการทำ association จะมี ไข่,ขนมปังฝรั่งเศล,น้ำส้มสายชู้,น้ำมันถั่วเหลืองออกมาเยอะ เพราะมีความถี่เยอะ

**ตอนที่ 2**

**Parameter setup**



ใช้SubsetByExpressionในการ filter ข้อมูลที่จำเป็นออกมาโดยเราต้อง set ค่าใน expression โดยที่ ATT หมายถึง เลือก attribute ที่เราต้องการ filter ข้อมูล ตัวอย่างเช่น ATT1 = 1 หมายความว่า ทำการเลือก attribute ที่ 1 ว่าจะนำค่าที่เป็น 1 ออกมา และattributeที่นำมา expression ต้องเป็นชนิด numeric

**Assosiation’s parameter**

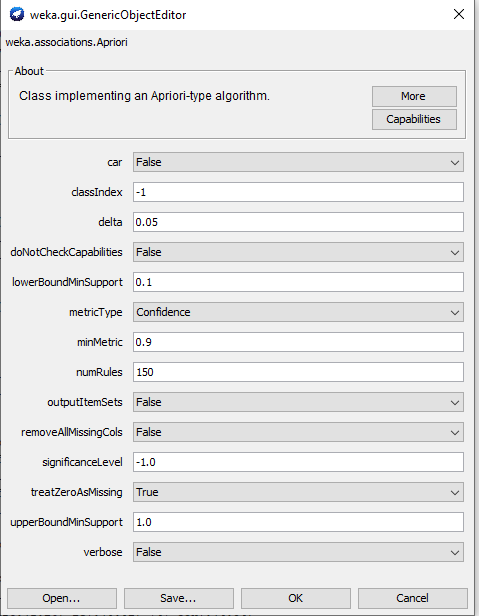
ค่าที่ได้หลังจาก association เเล้วนำมาพิจรณาคือ

ค่า confident คือค่าความมั่นใจที่เกิด LHS เเล้วจะเกิด RHS

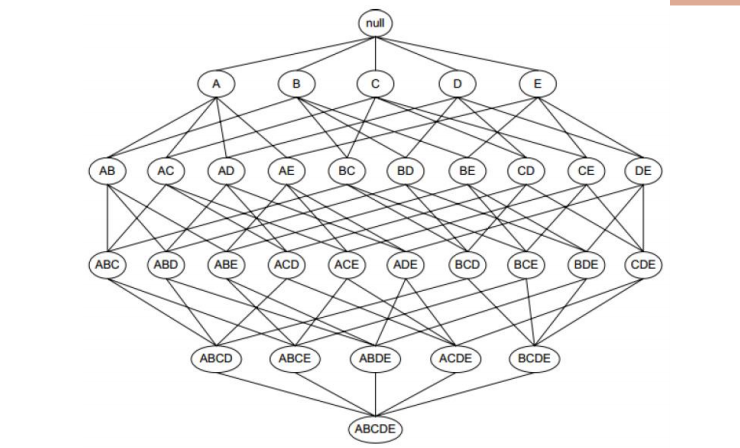
ค่า lift เเละ leverage คือค่าความเกี่ยวข้องของ LHS กับ RHS

ค่า conviction คือค่าความขัดเเย้งที่เกิดLHS เเต่จะไม่เกิด RHS

**การตั้งค่า Apriori**



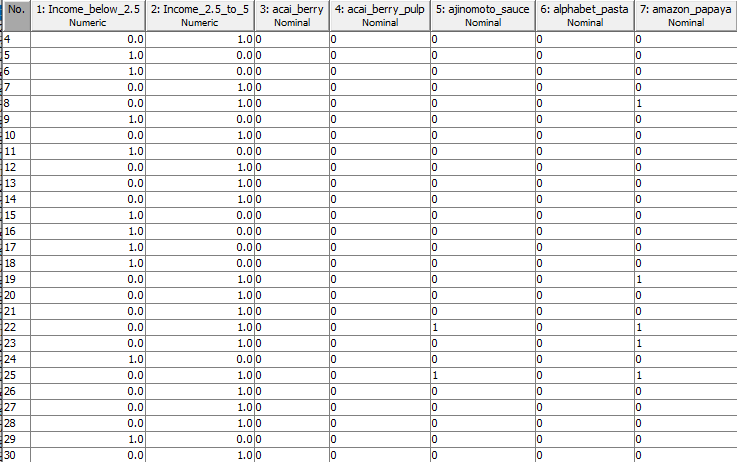
1.ทำการเซ็ทค่าตัว treatZeroAsMissing ใหม่จาก false เป็น true เพื่อลดการนำค่า 0 มาคิดเนื่องจากจะทำให้ค่า Apriori ออกมามีค่ามากเกินไป เพื่อการลด runtime เเละการใช้งาน ram เพราะการทำ Apriori จะใช้ algorithm เป็นเเบบ brute force ในการหาความสัมพันธ์ของ itemsets



2.เปลี่ยนค่า numRules ให้มีจำนวนมากขึ้นเพื่อที่เราจะมีค่า apriori มาใช้ในการวิเคราะห์ได้มากขึ้น

(หา Bestrule)

**ค่า Income ที่น้อยกว่า 5 หลังจากทำการ filter ข้อมูล**



ค่า best rule Income น้อยกว่า 5 ที่ทำการเลือกมา 3 ตัว ได้แก่

1. chocolate\_powder=1 popping\_corn=1 35 ==> egg=1 33

conf:(0.94) lift:(1.27) lev:(0.02) [7]

2. egg=1 garlic=1 milk\_in\_plastic\_bag\_category\_C=1 34 ==> french\_bread=1 32

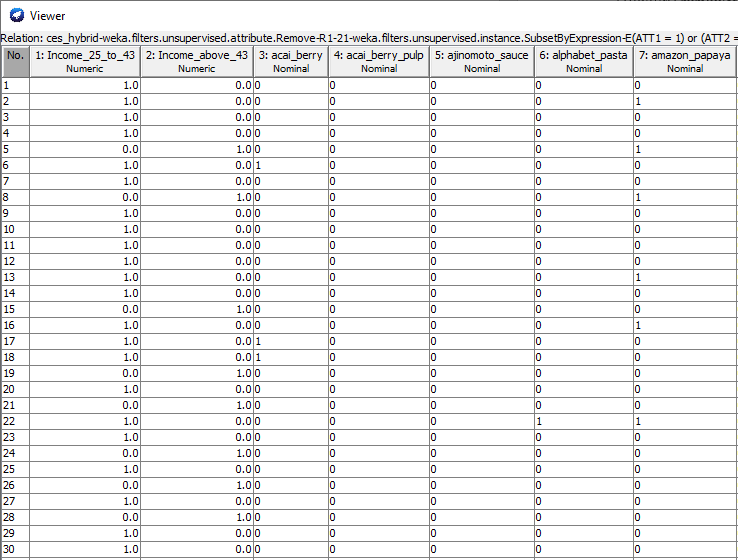
conf:(0.94) lift:(1.35) lev:(0.03) [8]

3. egg=1 garlic=1 milk\_in\_plastic\_bag\_category\_C=1 34 ==> soy\_oil=1 32

conf:(0.94) lift:(1.19) lev:(0.02) [5]

จาก best rule ที่ได้เลือกมานั้น เราสามารถทำนายได้ว่า ถ้าเลือกซื้อ ผงซ็อกโกเเล็ต เเละป็อปคอน จะต้องซื้อไข่ด้วย เเละ ถ้าเลือกซื้อ ไข่ กระเทียม เเละนมถุง จะต้องซื้อขนมปังฝรั่งเศล หรือน้ำมันถั่วเหลืองด้วย

**ค่า Income ที่มากกว่า 25 หลังจากทำการ filter ข้อมูล**



ค่า best rule Income มากกว่า 25 ที่ทำการเลือกมา 3 ตัว ได้แก่

1.egg=1 french\_bread=1 guava\_candy=1 45 ==> mayonnaise=1 42

conf:(0.93) lift:(1.32) lev:(0.04) [10]

2. egg=1 guava\_candy=1 mayonnaise=1 45 ==> french\_bread=1 42

conf:(0.93) lift:(1.29) lev:(0.04) [9]

3. egg=1 mayonnaise=1 whipped\_cream=1 46 ==> vinegar=1 42

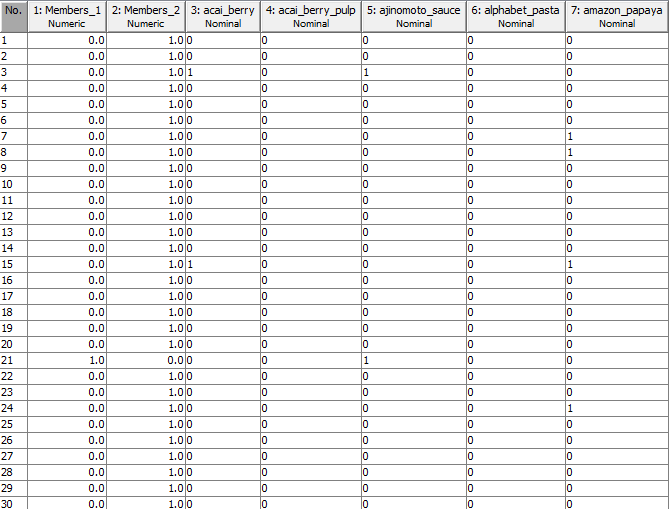
conf:(0.91) lift:(1.3) lev:(0.04) [9]

จาก best rule ที่ได้เลือกมานั้น เราสามารถทำนายได้ว่า เมื่อผู้มีรายได้มากว่า 25 ทำการซื้อ ไข่ ขนมปังฝรั่งเศส และลูกอมรสฝรั้ง จะทำการซื้อมายองเนสด้วย และ ถ้าทำการเลือกซื้อ ไข่ ลูกอมรสฝรั้ง และมายองเนส จะทำการซื้อขนมปังฝรั่งเศสด้วย และ ถ้าทำการซื้อ ไข่ มายองเนส และวิปครีม จะทำการซื้อน้ำส้มสายชูด้วย

สรุปจาก best rule ทั้งของ Income น้อยกว่า 5 และ Income มากกว่า 25 จะสังเกตได้ว่าทั้ง 2 กลุ่มจะนิยมซื้อไข่ และขนมปังฝรั่งเศส แต่ที่ต่างกันคือกลุ่มที่มี Income น้อยกว่า 5 จะนิยมเลือกซื้อน้ำมันถั่วเหลือง และกระเทียม ส่วนกลุ่มที่มี Income มากกว่า 25 จะนิยมเลือกซื้อน้ำส้มสายชู และมายองเนส

**ตอนที่ 3**

**ค่า Members <=2 หลังจากทำการ filter ข้อมูล**

****

ค่า best rule ค่า Members <=2 ที่ทำการเลือกมา 3 ตัว ได้แก่

1. chocolate\_powder=1 french\_bread=1 garlic=1 mayonnaise=1 34 ==> vinegar=1 33

conf:(0.97) lift:(1.54) lev:(0.04) [11]

2. chocolate\_powder=1 mayonnaise=1 vinegar=1 wheat\_flour\_special=1 33 ==> egg=1 32

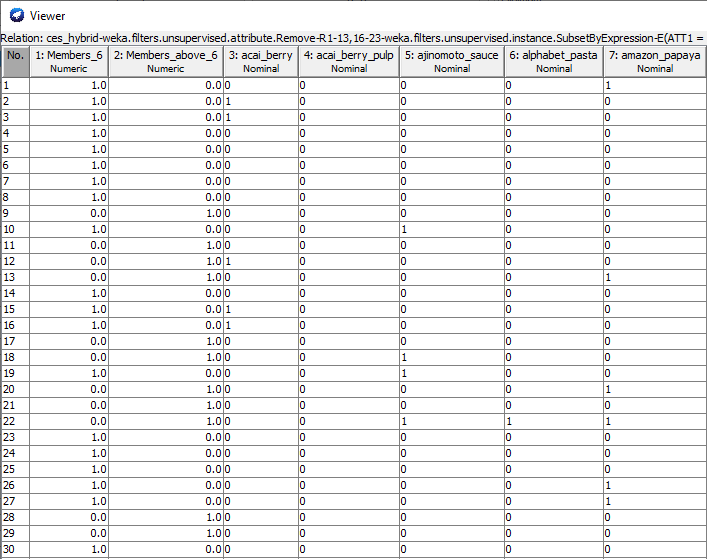
conf:(0.97) lift:(1.39) lev:(0.03) [8]

3. french\_bread=1 green\_beans=1 wheat\_flour\_special=1 37 ==> egg=1 36

conf:(0.97) lift:(1.39) lev:(0.03) [10]

จาก best rule ที่ได้เลือกมานั้น เราสามารถทำนายได้ว่า ถ้าเลือกซื้อ ผงซ็อกโกเเล็ต ขนมปังฝรั่งเศส หัวหอม และมายองเนส จะทำการซื้อน้ำส้มสายชูด้วย เเละถ้าเลือกซื้อ ผงซ็อกโกเเล็ต มายองเนส น้ำส้มสายชู เเละแป้งสาลีพิเศษ จะทำการซื้อไข่ด้วย เเละถ้าเลือกซื้อขนมปังฝรั่งเศส ถั่วเขียว และแป้งสาลีพิเศษ จะทำการซื้อไข่ด้วย

**ค่า Members >= 6 หลังจากทำการ filter ข้อมูล**



ค่า best rule ค่า Members >= 6 ที่ทำการเลือกมา 3 ตัว ได้แก่

1.egg=1 mayonnaise=1 popping\_corn=1 vinegar=1 47 ==> french\_bread=1 44

conf:(0.94) lift:(1.22) lev:(0.03) [8]

2.chocolate\_powder=1 french\_bread=1 garlic=1 mayonnaise=1 soy\_oil=1 46 ==> vinegar=1 44 conf:(0.96) lift:(1.32) lev:(0.04) [10]

3.chocolate\_powder=1 ketchup=1 soy\_oil=1 vinegar=1 42 ==> mayonnaise=1 39

conf:(0.93) lift:(1.52) lev:(0.05) [13]

จาก best rule ที่ได้เลือกมานั้น เราสามารถทำนายได้ว่า ถ้าเลือกซื้อ ไข่ ป็อปคอน และน้ำส้มสายชู จะทำการซื้อขนมปังฝรั่งเศสด้วยด้วย เเละถ้าเลือกซื้อ ผงซ็อกโกเเล็ต มายองเนส ขนมปังฝรั่งเศส และ  
หัวหอม จะทำการซื้อน้ำส้มสายชูด้วย เเละถ้าเลือกซื้อผงซ็อกโกเเล็ต ซอสมะเขือเทศ น้ำมันถั่วเหลือง และน้ำส้มสายชู จะทำการซื้อมายองเนสด้วย

สรุปจาก best rule ทั้งของ Members <=2 และ Members >= 6 จะสังเกตได้ว่าทั้ง 2 กลุ่มจะนิยมซื้อผงซ็อกโกเเล็ต ขนมปังฝรั่งเศส และหัวหอม ส่วนกลุ่มที่ Members >= 6 จะนิยมซื้อน้ำส้มสายชู และมายองเนส ส่วนกลุ่มที่ Members <=2 จะนิยมซื้อถั่วเขียว และแป้งข้าวสาลี

**ตอนที่ 4**

สรุปไม่ว่าจะมีรายได้ในช่วงไหน หรือจำนวนสมาชิกเท่าไหร่ก็จะนิยมซื้อขนมปังฝรั่งเศสกับไข่ ส่วนถ้าดูตามรายได้จะนิยมซื้อน้ำส้มสายชู ไวน์และมายองเนส ส่วนถ้าดูตามจำนวนสมาชิกในครอบครัวจะนิยมซื้อถั่วเขียว เนย โยเกิร์ตผลไม้ และแป้งข้าวสาลี

**อ้างอิง**

Gonçalves, E. C. (2014). A Human-Centered Approach for Mining Hybrid-Dimensional Association Rules. Proceedings of the 17th International Conference on Information Fusion, (FUSION 2014), Salamanca, Spain.