



# Fortune Cookies Model

## โดยอาศัยเทคนิค Association Rules

เสนอ

ดร.ชนชาติย์ ฤทธิบำรุง

จัดทำโดย

นาย ชัยรัตน์	ปัญญาคุณ	รหัสนักศึกษา 6110422026
นาย ชลัช	มงคลอิรภัทร์	รหัสนักศึกษา 6110422032
นาย อติเทพ	กิติธีระกุล	รหัสนักศึกษา 6110422047

รายงานนี้เป็นส่วนหนึ่งของวิชา การวิเคราะห์ธุรกิจและวิทยาการข้อมูลเบื้องต้น (BADS6001)

ระดับปริญญาโท ภาคพิเศษ ภาคการเรียนที่ 1 ประจำปีการศึกษา 2561

สาขาวิชาเอกวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

## ที่มาและเหตุผล

การวิเคราะห์ข้อมูลเพื่อค้นหาคุณลักษณะเด่นที่ทำให้สมาชิกในวง BNK48 ได้รับความสนใจจากเหล่าแฟนคลับ ด้วยเทคนิคการหาความสัมพันธ์ (Association Rules)

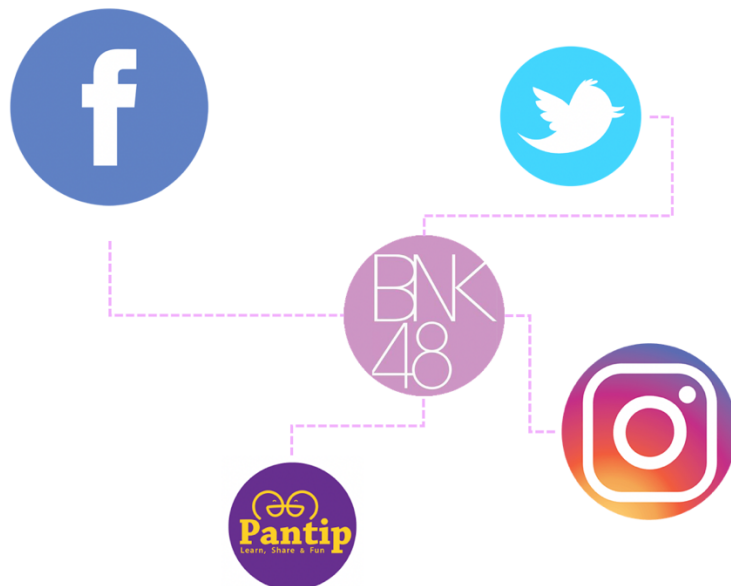
เนื่องจากกระแสของวง BNK48 ที่กำลังโด่งดังในปัจจุบันนี้มีมาจากหลายปัจจัย ไม่ว่าจะเป็นด้วยบุคลิก ท่าทาง รูปร่าง หน้าตา สัดส่วน อุปนิสัย หรือแม้แต่ช่วงอายุที่แตกต่างกันออกไปของสมาชิก ส่งผลให้สมาชิกแต่ละคนได้รับความนิยมที่แตกต่างกัน รายงานฉบับนี้ต้องการศึกษาวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ของบุคลิกและหน้าตาของสมาชิกวง BNK48 ที่คาดว่า การมีองค์ประกอบเหล่านี้อยู่รวมกันแล้วจะทำให้ได้รับความนิยมจากแฟนคลับ



รูปที่ 1

## แนวทางการเก็บข้อมูล

แหล่งข้อมูลที่เกี่ยวข้องกับสมาชิกวง BNK48 นั้นมีอยู่ในโลกออนไลน์อยู่มากมาย มีทั้งข้อมูลที่เป็นทางการจากต้นสังกัด และยังมีข้อมูลจากกลุ่มผู้สนับสนุนหรือที่เรียกว่า “โอตะ” เพียงแต่ข้อมูลเหล่านั้นนั้นกระจัดกระจายไม่ได้ถูกจัดเก็บหรือรวบรวมให้อยู่ในโครงสร้างที่พร้อมใช้งานสำหรับการวิเคราะห์ข้อมูล



รูปที่ 2

โดยเมื่อตั้งวัตถุประสงค์เสร็จ จึงเริ่มทำการออกแบบตัวแปรต่างๆ สำหรับนำเข้าโมเดล และเริ่มเก็บข้อมูลจากแหล่งข้อมูล ปัญหาหลังจากเริ่มเก็บข้อมูลพบว่า ข้อมูลที่ต้องการไม่มีการพูดถึงอย่างชัดเจน แต่เนื่องจากมีแหล่งที่สามารถเก็บข้อมูลได้ชัดเจน นั่นคือกลุ่มผู้สนับสนุนหรือที่เรียกว่า “โอตะ” ของ วง BNK48 โดยมีวิธีและขั้นตอนในการเก็บรวบรวมข้อมูล เป็นไปตามขั้นตอนดังต่อไปนี้

1. เก็บข้อมูลความนิยมของสมาชิกจากเฟสบุ๊กแฟนเพจของสมาชิกวง BNK48 ทั้งหมดด้วย Web Scrapping ภาษา Python และ Library BeautifulSoup
2. เลือกสมาชิก 7 อันดับ โดย Ranking จากยอด Like เฟสบุ๊คสูงสุด เพื่อกำหนดขอบเขตข้อมูลสำหรับการเก็บข้อมูล และวิเคราะห์ผล ได้สมาชิก 7 อันดับ ประกอบด้วย
  - 2.1 เฉอปราง - เฉอปราง อารีย์กุล
  - 2.2 ป๊อ - ป๊อสิริณ คิยะกร
  - 2.3 อร - พัชชนันท์ เจียจิรโชติ
  - 2.4 เจนนิชฐ - เจนนิชฐ โอประเสริฐ
  - 2.5 เนย - กานต์ธีรา วัชรทัศนกุล
  - 2.6 โมบายล์ - พิมรภัศ ผดุงวัฒนะโชค
  - 2.7 มิวสิค - แพรวา สุธรรมพงษ์
3. กำหนดตัวแปรสำหรับใช้ในแบบสอบถาม จากวัตถุประสงค์เพื่อวัดความชอบ/ความพึงพอใจของสมาชิกในวงจากหน้าตา ตัดส่วน อุปนิสัย โดยตัวแปรที่ใช้ในแบบสอบถาม ประกอบด้วย
  - 3.1 ความพึงพอใจใน โครงหน้า, หู, ตา, จมูก, ปาก, คาง, คิ้ว, สัดส่วน, ส่วนสูง, น้ำหนัก, รอบเอว
  - 3.2 ความพึงพอใจในอุปนิสัย ประกอบด้วย ร่าเริง, ซื่ออ่อน, จี๊ดๆ, แข็งแกร่ง, เป็นผู้นำ, แมน, บอบบาง, น่ารัก, เซ็กซี่, หัว
4. จัดทำแบบสอบถามโดยคำถามที่เกี่ยวข้องกับประกอบของหน้าตา จัดให้อยู่ในรูปแบบของการให้คะแนนตามความชื่นชอบ (Rank 1-5) และคำถามที่เกี่ยวกับบุคลิกภาพ จัดให้อยู่ในรูปแบบเลือกได้เพียง 1 ข้อ
5. แจกแบบสอบถาม ที่สร้างด้วย Google Form ลงใน BNK48 Community ต่างๆ (<https://goo.gl/uZeEzV>)

## วิธีการวิเคราะห์ข้อมูล

หลังจากได้รับข้อมูลจากแบบสอบถาม จะพบว่าข้อมูลนั้นยังไม่สามารถนำไปใช้งานได้ทันที จึงจำเป็นต้องผ่านกระบวนการในการจัดการข้อมูลให้พร้อมใช้สำหรับการวิเคราะห์ โดยมีขั้นตอนดังต่อไปนี้

1. จัดเตรียมข้อมูลให้อยู่ในรูปแบบที่ต้องการ โดยการจัดการข้อมูลที่น่าสนใจให้อยู่ในรูปแบบของแถว

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	lineament	ear	eye	nose	mouth	chin	eyebrow	outline	height	Waistline	weight	Cherprang [lineament]	Pun [lineament]	Orn [lineament]
2		5	4	5	3	3	4	3	5	5	3		1	5
3		1	3	3	3	2	1	5	1	2	1		4	1
4		5	3	5	1	3	2	4	1	4	1		3	5
5		5	2	4	4	2	2	1	1	3	4		4	5
6		2	1	5	4	4	2	2	1	5	1		2	5
7		1	3	1	4	4	5	5	5	5	4		5	3
8		5	3	3	1	2	3	4	1	3	1		1	5
9		2	3	5	1	4	4	5	2	5	4		1	1
10		1	1	3	1	3	3	1	5	5	3		4	5
11		1	5	3	1	4	5	4	4	4	2		2	5
12		5	2	5	2	2	5	2	3	5	3		1	5
13		2	5	1	5	4	5	1	5	3	1		4	5
14		3	4	3	5	3	1	1	5	1	4		5	5
15		1	5	1	2	3	2	4	4	1	5		5	2
16		4	3	4	3	1	3	4	5	4	3		2	5

รูปที่ 3

2. จัด Ranking เลือก 7 อันดับที่มียอด Like สูงสุดมาทำ Profile Profiling เพื่อนำตัวแปรไปใช้หาความสัมพันธ์ โดยการแยกตัวแปรออกเป็น 2 ประเภท คือ
  - 2.1 ทางกายภาพ เช่น หู, ตา, จมูก, ปาก
  - 2.2 ทางบุคลิกภาพ เช่น อุปนิสัย
3. รวบรวมคะแนนที่ได้จากการโหวต 1-5
  - 3.1 องค์ประกอบที่เรียงลำดับมีคะแนนรวมสูงสุด

	VoteAll	Vote5
Orn [eye]	2264	1320
Mobile [lineament]	2248	1240
Cherprang [mouth]	2246	1230
Orn [nose]	2059	1085
Cherprang [eyebrow]	1554	590
Mobile [ear]	1543	580
Noey [chin]	1506	485

รูปที่ 4

- 3.2 องค์ประกอบที่เรียงลำดับมีคะแนนโหวต 5 สูงสุด

	VoteAll	Vote5
Cherprang [eye]	2221	1880
Pun [lineament]	2201	1875
Jennis [mouth]	2185	1845
Orn [nose]	2059	1085
Cherprang [eyebrow]	1554	590
Mobile [ear]	1543	580
Mobile [chin]	1477	500

รูปที่ 5

จากการรวบรวมคะแนนพบว่า รายการที่มีคะแนนรวมสูงสุดกับรายการที่ได้รับการโหวตคะแนน 5 สูงสุด ในแต่ละองค์ประกอบนั้นเป็นคนละรายการกัน

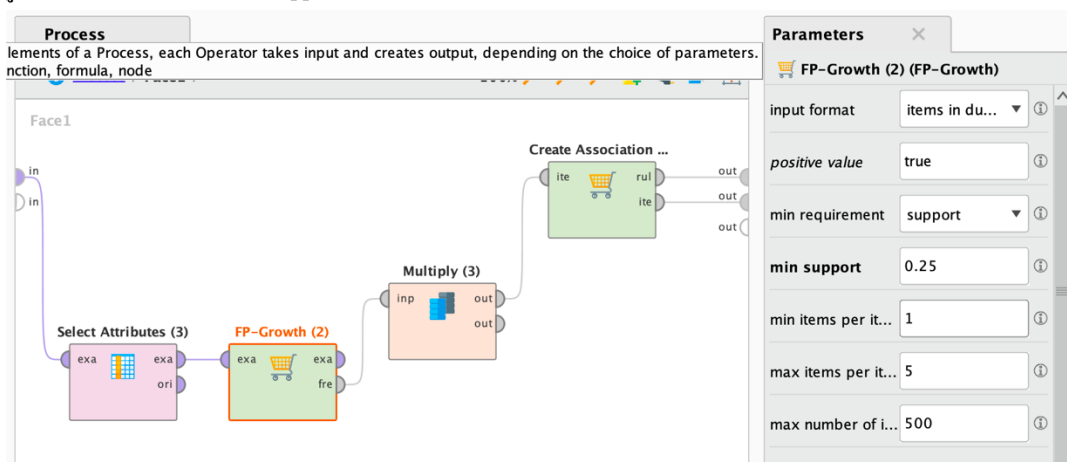
- 4 แปลงข้อมูล Numeric ให้อยู่ในรูปแบบของ Binominal ด้วยโปรแกรม Rapid Miner โดยพิจารณาจากองค์ประกอบที่ได้รับการโหวตคะแนน 5 แปลงค่าเป็น True และองค์ประกอบที่ได้รับการโหวตคะแนน 1-4 แปลงค่าเป็น False

ExampleSet (500 examples, 0 special attributes, 94 regular attributes)

Row No.	lineament	ear	eye	nose	mouth	chin	eyebrow
1	true	false	true	false	false	false	false
2	false	false	false	false	false	false	true
3	true	false	true	false	false	false	false
4	true	false	false	false	false	false	false
5	false	false	true	false	false	false	false
6	false	false	false	false	false	true	true
7	true	false	false	false	false	false	false
8	false	false	true	false	false	false	true
9	false	false	false	false	false	false	false
10	false	true	false	false	false	true	false

รูปที่ 6

- 5 หาความสัมพันธ์ของข้อมูลที่เกิดขึ้นซ้ำกัน (Frequency Itemset) ด้วย FP-Growth Algorithm ปรับค่า Support ที่เหมาะสมกับข้อมูล ณ การวิเคราะห์นี้ใช้ค่า Support = 0.25

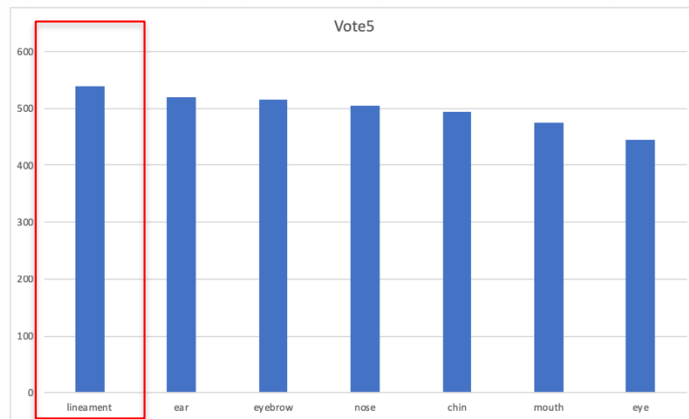


รูปที่ 7

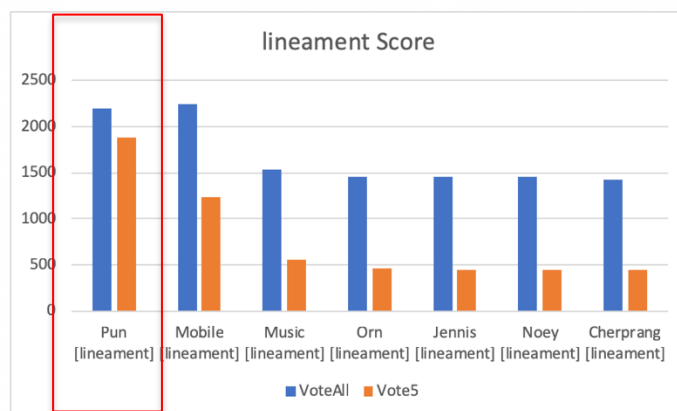
No. of Sets: 500 Total Max. Size: 5						
Min. Size: 1	Size ↑	Support	Item 1	Item 2	Item 3	Item 4
Max. Size: 5	4	0.360	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Orn [nose]
Contains Item:	4	0.226	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Pun [eye]
Update View	4	0.158	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Noey [eye]
	4	0.156	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Music [eye]
	4	0.140	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Music [lineament]
	4	0.134	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Jennis [eye]
	4	0.136	Cherprang [eye]	Pun [lineament]	Jennis [mouth]	Mobile [nose]

รูปที่ 8

- 6 นำข้อมูลที่เกิดซ้ำกัน(Frequency Itemset) มาสร้างกฎความสัมพันธ์ (Association Rules) ด้วยโปรแกรม Rapid Miner จากข้อมูลการโหวต องค์ประกอบที่สนใจ พบว่าจะเน้นในส่วนของ โครงหน้าสูงสุด และ ผลโหวตองค์ประกอบ 5 ของ โครงหน้าสูงสุด เป็นของ สมาชิกชื่อบัญ



รูปที่ 9



รูปที่ 10

- 7 เมื่อนำข้อมูลคะแนนในส่วนของ โครงหน้าสูงสุดและผลโหวตองค์ประกอบ 5 ของโครงหน้าสูงสุด พบว่าเป็นของ สมาชิกชื่อบัญ มาร่วมพิจารณาในการเลือกกฎความสัมพันธ์ที่เกิดขึ้น พบว่าจากกฎความสัมพันธ์ที่เกิดขึ้นจากโมเดล จะเลือก กฎความสัมพันธ์ที่ 1653 ขึ้นมาอธิบาย

No.	Premises	Conclusion	Supp... ↓	Lift	Confide...
1653	Pun [lineament]	Cherprang [eye], Jennis [mouth], Orn [nose]	0.360	1.326	0.480

รูปที่ 11

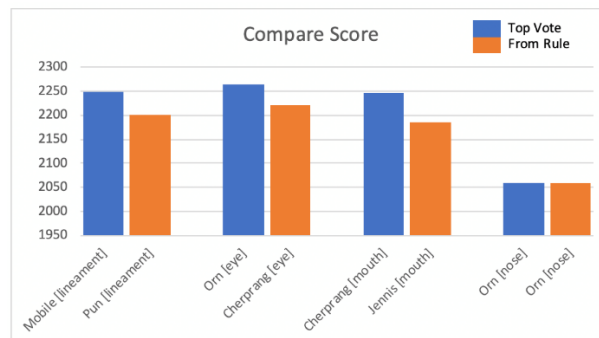
ค่า Support มีค่าเท่ากับ 0.360 นั่นคือ Rules นี้มีโอกาสเกิดขึ้นได้ถึง 36% จากแบบสอบถามทั้งหมด 500 คน มีผู้ตอบแบบสอบถามมากถึง 180 คน ที่ให้ 5 คะแนน กับ โครงหน้าของปัญญา, ตาของเฉอปร่าง, ปากของเจนนิส และ จมูกของอร

ค่า Confident มีค่าเท่ากับ 0.480 นั่นคือ ในบรรดาผู้ที่ชื่นชอบโครงหน้าของปัญญา มีถึง 48% ที่ชอบ ตาของเฉอปร่าง, ปากของเจนนิส และ จมูกของอร

ค่า Lift คือ มีค่าเท่ากับ 1.326 นั่นคือ Rules ดังกล่าวเกิดขึ้นมากกว่าที่คาดการณ์ไว้ถึง 1.326 เท่า กรณีที่ค่า Lift > 1 หมายความว่า LHS และ RHS แปรผันตามกัน และมีศักยภาพมากเพียงพอที่จะเป็น Rules ที่ต้องพิจารณา

## ผลลัพธ์ที่ได้และการนำไปใช้ประโยชน์

จากการวิเคราะห์ข้อมูลสามารถอธิบายได้ว่า ใบหน้าที่ประกอบไปด้วย โครงหน้าของปัญญา (Pun [lineament]), ตาของเฉอปร่าง (Cherprang [eye]), ปากของเจนนิส (Jennis [mouth]) และ จมูกของอร (Orn [nose]) คือองค์ประกอบที่กลุ่มตัวอย่างในการสำรวจมีแนวโน้มที่จะให้ Member ของวง BNK48 มีหน้าตาที่ประกอบไปด้วยองค์ประกอบดังกล่าว เมื่อมีองค์ประกอบมากกว่า 1 อย่างแล้วเราไม่สามารถนำข้อมูลที่มีคะแนนสูงสุดมารวมหรือจับคู่กันได้เลยทันที เนื่องจากเมื่อมีองค์ประกอบที่มากกว่า 1 อย่างอยู่ร่วมกัน ความสัมพันธ์ที่เกิดขึ้นอาจจะไม่ได้เป็นไปตามคะแนนสูงสุดที่ได้ เช่น จากข้อมูล พบว่า องค์ประกอบที่ได้รับคะแนนสูงสุดในแต่ละองค์ประกอบ โครงหน้าของโมบาย (Mobile [lineament]), ตาของอร (Orn [eye]), ปากของเฉอปร่าง (Cherprang [mouth]) และ จมูกของอร (Orn [nose]) แต่การสร้างกฎความสัมพันธ์กลับแสดงข้อมูลที่บอกว่า โครงหน้าของปัญญา (Pun [lineament]), ตาของเฉอปร่าง (Cherprang [eye]), ปากของเจนนิส (Jennis [mouth]) และ จมูกของอร (Orn [nose]) คือองค์ประกอบที่มีคนสนใจร่วมกันมากกว่า



รูปที่ 12

การวิเคราะห์ข้อมูลด้วยกฎความสัมพันธ์นั้นจะช่วยทำนายเกี่ยวกับความสัมพันธ์ของการเกิดเหตุการณ์ร่วมกัน ตั้งแต่ 2 อย่างขึ้นไปในกรณีนี้เราใช้ทำนายองค์ประกอบของหน้าตาของสมาชิกคนใหม่ ที่คาดว่าจะได้รับความสนใจจากรจะประกอบไปด้วยอะไรบ้าง เพื่อใช้ในการเป็นไกด์ไลน์ในการเฟ้นหาสมาชิกวงรุ่นถัดไป