

# การวิเคราะห์สารอาหารเพื่อหลีกเลี่ยงโรคเบาหวาน โดยใช้ขั้นตอนการแบ่งกลุ่มข้อมูล Nutrient Analysis to Avoid Diabetes by Using Data Clustering Algorithms

ถาวรีย สุขเกษม (Thawaree Sukkasem)\* อุรีรัฐ สุขสวัสดิ์ชน (Ureerat Suksawatchon)\*\*  
และ จักริน สุขสวัสดิ์ชน (Jakkarin Suksawatchon)\*\*

## บทคัดย่อ

งานวิจัยนี้เป็นงานวิจัยเชิงทดลองเพื่อหาวิธีการแบ่งกลุ่มและผลการแบ่งกลุ่มที่เหมาะสมที่สุด เพื่อใช้ในการวิเคราะห์สารอาหารในแต่ละประเภทอาหารของผู้ป่วยโรคเบาหวาน โดยวิธีการแบ่งกลุ่มที่ใช้ ได้แก่ การแบ่งกลุ่มแบบเคมีนส์แบบเคมีนส์พลัสพลัส แบบเคมีดอยด์ และแบบเอ็กซ์มีนส์ ข้อมูลรายการอาหารที่ใช้สำหรับการแบ่งกลุ่มมาจากฐานข้อมูลรายการอาหารของสถาบันโภชนาการ มหาวิทยาลัยมหิดลจำนวน 1,741 รายการ โดยที่แต่ละรายการอาหารประกอบไปด้วยสารอาหารที่เกี่ยวข้องกับโรคเบาหวาน 8 ชนิด และเมื่อพิจารณาจากการวัดประสิทธิภาพการแบ่งกลุ่มพบว่าผลการแบ่งกลุ่มที่เหมาะสมที่สุดของรายการอาหารประเภทอาหารเช้าและเครื่องดื่ม ได้จากการแบ่งกลุ่มด้วยวิธีเคมีดอยด์ ประเภทขนมและผลไม้ได้จากการแบ่งกลุ่มด้วยวิธีเอ็กซ์มีนส์ และประเภทอาหารจานหลักและอาหารว่างได้จากการแบ่งกลุ่มด้วยวิธีเคมีนส์พลัสพลัส

**คำสำคัญ:** การแบ่งกลุ่มข้อมูล การวิเคราะห์สารอาหาร โรคเบาหวาน

## Abstract

This paper is the experimental research to find the most appropriate clustering algorithm and clustering result for each food categories. The clustered foods are used to analyze the nutrient for Diabetes person. The clustering algorithm used in this work are K-means, K-means++, K-medoids and

X-means. This study is based on the databases provided from Institute of Nutrition, Mahidol University, Thailand, including 1,741 meals with 8 related nutrients. When considering with cluster validity, we found that the suitable clustering results of breakfasts and beverages categories are obtained from K-medoids. The appropriated clustering result for desserts and fruits is obtained from X-means and for main dishes and snacks groups are obtained from K-means++.

**Keywords:** Data Clustering, Nutrient Analysis, Diabetes.

## 1. บทนำ

โรคเบาหวานเป็นหนึ่งในโรคไม่ติดต่อเรื้อรังที่สำคัญที่มีอัตราการตายเพิ่มสูงขึ้นอย่างต่อเนื่อง ซึ่งถือได้ว่าเป็นโรคที่คร่าชีวิตผู้คนมากเป็นอันดับต้นๆ อีกทั้งยังเป็นปัจจัยที่เพิ่มความเสี่ยงของการเกิดโรคแทรกซ้อนอื่นๆ ได้อีกด้วย วิธีการดูแลสุขภาพผู้ที่ป่วยเป็นโรคเบาหวานนอกจากจะต้องทานยาตามคำแนะนำของแพทย์อย่างเคร่งครัดแล้ว “การควบคุมการรับประทานอาหาร” ก็เป็นอีกหนึ่งปัจจัยสำคัญในการรักษาโรคเบาหวานด้วย ผู้ป่วยควรเลือกรับประทานอาหารให้ครบหมวดหมู่และมีความหลากหลาย อีกทั้งควรเลือกรับประทานผักผลไม้ที่มีเส้นใยอาหารมาก หลีกเลี่ยงอาหารที่หวานจัดและไขมันสูง และต้องคำนึงถึงปริมาณพลังงานของอาหารไม่ให้เกินกว่าที่ร่างกายต้องการในหนึ่งวัน [1] ซึ่งสิ่งเหล่านี้ นักโภชนาการจะต้องให้ความรู้ และทำความเข้าใจกับผู้ป่วยโรคเบาหวานอย่างละเอียด

\* คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

\* Faculty of Informatics, Burapha University.

\*\* ห้องปฏิบัติการวิจัย Mobile Application Developers Incubation

\*\* Mobile Application Developers Incubation.

แต่เนื่องจากรายการอาหารหนึ่งๆ ประกอบไปด้วยสารอาหารหลายอย่างที่มีผู้ป่วยเบาหวานต้องพิจารณาในการเลือกรับประทานอาหาร ทำให้ยากต่อการทำความเข้าใจและนำไปใช้ในชีวิตประจำวัน ดังนั้นหากผู้ป่วยโรคเบาหวานสามารถทราบรายการอาหารใดบ้างที่สามารถรับประทานได้ รายการอาหารใดบ้างที่ควรหลีกเลี่ยงหรือมีรายการอาหารอื่นใดอีกบ้างที่มีสารอาหารคล้ายกันให้สามารถเลือกรับประทานได้ โดยมีเครื่องมือช่วยแบ่งกลุ่มรายการอาหารให้กับผู้ป่วย ก็น่าจะช่วยให้ผู้ป่วยสามารถเลือกรับประทานอาหารที่มีประโยชน์และหลีกเลี่ยงอาหารที่อาจเพิ่มความรุนแรงของโรคเบาหวานได้

จากการศึกษางานวิจัยที่เกี่ยวข้องมีหลายงานวิจัยที่นำเทคนิคการแบ่งกลุ่มมาใช้ในการแบ่งกลุ่มรายการอาหาร โดยแบ่งออกเป็น 2 วิธีใหญ่ๆ ได้แก่ การแบ่งกลุ่มแบบลำดับชั้น (Hierarchical clustering) และการแบ่งกลุ่มแบบแบ่งกลุ่ม (Partition clustering) ตัวอย่างงานวิจัยที่ใช้การแบ่งกลุ่มแบบลำดับชั้น เช่น งานวิจัยของ Huan-Chung Li และ Wei-Min Ko [2] ได้นำเสนอวิธีการสร้างออนโทโลยีอาหารอัตโนมัติสำหรับการดูแลผู้ป่วยเบาหวาน โดยใช้เทคนิคการแบ่งกลุ่มแบบลำดับชั้น (Hierarchical Clustering Algorithm : HCA) ด้วยวิธีรวมกลุ่มแบบเฉลี่ย (Average) และการจัดลำดับค่าคงที่ในการสร้างกลุ่มของอาหารเพื่อจัดลำดับอาหารที่จะแนะนำตามความคล้ายกันของระดับสารอาหาร แต่เนื่องจากวิธีการจัดกลุ่มแบบ HCA นี้เมื่อแบ่งกลุ่มข้อมูลแล้วจะไม่มี การคำนวณซ้ำเพื่อปรับปรุงการจัดกลุ่มในขั้นตอนก่อนหน้าได้ อีกทั้งต้องกำหนดจำนวนกลุ่มข้อมูลที่ต้องการ ซึ่งผู้ใช้ต้องพิจารณาเองจากโครงสร้างต้นไม้หรือเดนโดแกรม (Dendrogram)

ในงานวิจัยของ Phanich และคณะ [3] ได้เสนอระบบการให้คำแนะนำอาหารโดยใช้การวิเคราะห์ด้วยการแบ่งกลุ่มสำหรับผู้ป่วยโรคเบาหวานด้วยวิธีเคมีนส์ (K-means Clustering) ร่วมกับแผนผังการจัดระเบียบตัวเอง (Self-Organization Map, SOM) โดยแบ่งข้อมูลรายการอาหารตามลักษณะของสารอาหารได้จำนวน 22 กลุ่มและแบ่งตามระดับสารอาหารที่มีผลต่อโรคเบาหวานจำนวน 8 สารอาหารได้แก่ พลังงาน คาร์โบไฮเดรต ไขมัน โปรตีน โยอาหาร วิตามินบีหนึ่ง (ไทอะมิน) วิตามินซีและวิตามินอี และงานวิจัยของ Yanisa และคณะ [4] นำเสนอระบบให้คำปรึกษาด้านโภชนาการและ

วางแผนรายการอาหารสำหรับบุคคลทั่วไป โดยแบ่งข้อมูลรายการอาหารออกเป็น 5 ประเภท ได้แก่ อาหารเช้า (Breakfast) เครื่องดื่ม (Drink) อาหารจานหลัก (Main dish) ขนมและผลไม้ (Dessert and Fruit) และอาหารว่าง (Snack) ซึ่งสารอาหารที่ใช้พิจารณาในแต่ละประเภทได้แก่ พลังงาน คาร์โบไฮเดรต โปรตีนและไขมัน ส่วนเทคนิคในการแบ่งกลุ่มรายการอาหารในแต่ละประเภทใช้วิธีการแบ่งกลุ่มแบบเคมีนส์ (K-means Clustering) จากนั้นจึงสร้างรายการอาหารที่จะแนะนำในแต่ละมื้อด้วยฐานกฎ (Rule-based) พร้อมทั้งคำนวณปริมาณพลังงานและสารอาหารของผู้ใช้แต่ละคนด้วย เพศ อายุ น้ำหนัก ส่วนสูงและระดับของกิจกรรม

จากสองงานวิจัยที่กล่าวมาข้างต้นเป็นงานวิจัยที่ใช้วิธีการแบ่งกลุ่มรายการอาหารด้วยวิธีการแบบเคมีนส์ แม้ว่าจะเป็นวิธีที่นิยมและสามารถนำไปพัฒนาได้ง่าย แต่วิธีการแบบเคมีนส์นี้ผู้ใช้จะต้องกำหนดจำนวนกลุ่มที่ต้องการแบ่งล่วงหน้า ซึ่งหากผู้ใช้ไม่มีความรู้เกี่ยวกับลักษณะของข้อมูลมาก่อน จะไม่สามารถบอกได้ว่าจำนวนกลุ่มที่เหมาะสมมีจำนวนกลุ่มเท่าไร อีกทั้งวิธีการแบบเคมีนส์เริ่มต้นจากการกำหนดจุดศูนย์กลางกลุ่มด้วยการสุ่มเลือกจากข้อมูลตามจำนวนกลุ่มที่กำหนด ซึ่งอาจจะสุ่มได้จุดศูนย์กลางกลุ่มเริ่มต้นที่ไม่เหมาะสม จึงส่งผลให้ได้กลุ่มของรายการอาหารที่ไม่เหมาะสมไปด้วย นอกจากวิธีการแบบเคมีนส์แล้ว ยังมีวิธีการแบ่งกลุ่มวิธีการอื่นที่น่าจะเหมาะสมกว่าวิธีการแบบเคมีนส์

ดังนั้นงานวิจัยนี้จึงนำเสนองานวิจัยเชิงทดลอง โดยมีวัตถุประสงค์เพื่อหาวิธีการแบ่งกลุ่มที่เหมาะสมกับการแบ่งกลุ่มข้อมูลรายการอาหารตามสารอาหารที่มีความคล้ายคลึงกัน เพื่อให้ผู้ป่วยเบาหวานสามารถทราบรายการอาหารใดบ้างที่สามารถรับประทานได้ รายการอาหารใดบ้างที่ควรหลีกเลี่ยง โดยทดลองใช้วิธีการแบ่งกลุ่ม 4 วิธีที่ได้แก่ 1) วิธีการแบ่งกลุ่มแบบเคมีนส์ 2) แบบเคมีนส์พลัสพลัส (K-means++) 3) แบบเคมีดอยด์ (K-medoids) ซึ่งทั้งสองวิธีดังกล่าวนี้ช่วยแก้ปัญหาในเรื่องการเลือกจุดศูนย์กลางกลุ่มเริ่มต้น และ 4) วิธีการจัดกลุ่มเอ็กซ์มีนส์ (X-means) ที่แก้ปัญหาในเรื่องการกำหนดจำนวนกลุ่มที่เหมาะสมได้ จากนั้นใช้การวัดประสิทธิภาพการแบ่งกลุ่ม 2 วิธีได้แก่ Davies-Bouldin Index (DB) และ R-square Index (RS) เพื่อเป็นตัวตัดสินใจเลือกวิธีการที่เหมาะสมที่สุดจาก 4 วิธีการดังกล่าวข้างต้น เมื่อได้วิธีการและจำนวนกลุ่มที่

เหมาะสมแล้ว ก็จะนำผลการแบ่งกลุ่มที่ได้ไปวิเคราะห์สารอาหารของแต่ละกลุ่มต่อไป

## 2. วรณกรรมที่เกี่ยวข้อง

กำหนดให้  $X = \{X_1, X_2, \dots, X_n\}$  เป็นเซตของข้อมูลทั้งหมดจำนวน  $n$  ตัวอย่าง และข้อมูลหนึ่งตัวอย่าง  $X_i = \{X_{i1}, X_{i2}, \dots, X_{id}\}$  ประกอบด้วยปัจจัยที่ต้องการศึกษาจำนวน  $d$  ปัจจัย และให้  $C = \{c_1, c_2, \dots, c_k\}$  เป็นเซตของจุดศูนย์กลางกลุ่มข้อมูลทั้งหมด  $k$  กลุ่ม

### 2.1 ขั้นตอนการแบ่งกลุ่มแบบเคมีนส์ (K-means)

ในปี ค.ศ. 1967 James [5] ได้นำเสนอวิธีการแบ่งกลุ่มที่เรียกว่า เคมีนส์ ซึ่งเป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) ที่สามารถเข้าใจได้ง่าย โดยจะแบ่งกลุ่มข้อมูลที่มีลักษณะของข้อมูลที่คล้ายกันรวมไว้ด้วยกันเป็นกลุ่มๆ โดยลักษณะความคล้ายวัดจากระยะห่างแบบยูคลิด (Euclidean distance) จากจุดศูนย์กลางกลุ่มไปยังข้อมูลแต่ละตัว โดยมีขั้นตอนดังนี้

1. กำหนดจุดศูนย์กลางกลุ่มเริ่มต้นแบบสุ่มจำนวน  $k$  กลุ่ม
2. จัดข้อมูลเป็นกลุ่มด้วยการหาระยะทางระหว่างข้อมูลกับจุดศูนย์กลางกลุ่ม จัดให้ข้อมูลอยู่กลุ่มที่ข้อมูลมีระยะห่างจากจุดศูนย์กลางนั้นน้อยที่สุด ด้วยมาตรวัดแบบยูคลิด ดังสมการ (1)

$$dist(X_i, X_j) = \sqrt{\sum_{l=1}^d (x_{il} - x_{jl})^2} \quad (1)$$

โดยที่

$dist(X_i, X_j)$  คือ ระยะห่างระหว่างข้อมูลที่  $i$  กับ  $j$

$X_{il}$  คือ ข้อมูลที่  $i$  ปัจจัยที่  $l$

$X_{jl}$  คือ ข้อมูลที่  $j$  ปัจจัยที่  $l$

$d$  คือ จำนวนปัจจัยทั้งหมด

3. ปรับปรุงจุดศูนย์กลางกลุ่มใหม่ด้วยการหาค่าเฉลี่ยจากข้อมูลทุกตัวภายในกลุ่ม

4. ทำซ้ำตั้งแต่ข้อ 2. จนกระทั่งค่าเฉลี่ยของจุดศูนย์กลางแต่ละกลุ่มไม่มีการเปลี่ยนแปลง

### 2.2 ขั้นตอนการแบ่งกลุ่มแบบเคมีนส์พลัสพลัส (K-means++)

David และ Sergei [6] ได้นำเสนอการแบ่งกลุ่มแบบเคมีนส์พลัสพลัส เพื่อแก้ไขปัญหาจากการกำหนดจุดศูนย์กลางกลุ่มเริ่มต้นของวิธีการแบ่งกลุ่มแบบเคมีนส์

โดยแนวคิดของวิธีการนี้จะแบ่งออกเป็น 2 ส่วน ส่วนแรกคือการเลือกจุดศูนย์กลางกลุ่มกลุ่มแรกมาก่อน ด้วยวิธีการสุ่มจากเซตของข้อมูลมา 1 ตัวอย่าง จากนั้นจุดศูนย์กลางกลุ่มเริ่มต้นกลุ่มอื่นจะถูกเลือกด้วยค่าความน่าจะเป็นถ่วงน้ำหนักซึ่งคำนวณจากสัดส่วนของระยะห่างของข้อมูลแต่ละตัวไปยังจุดศูนย์กลางกลุ่มที่ใกล้ที่สุดที่เลือกไว้ก่อนหน้านี้กับผลรวมของระยะห่างของข้อมูลแต่ละตัวไปยังจุดศูนย์กลางที่ใกล้ที่สุดที่เลือกไว้ก่อนหน้านี้ เมื่อได้จุดศูนย์กลางกลุ่มเริ่มต้นจากขั้นตอนก่อนหน้านี้เรียบร้อยแล้ว ในส่วนที่สองเป็นการดำเนินการแบ่งกลุ่มเหมือนกับขั้นตอนการแบ่งกลุ่มแบบเคมีนส์ ซึ่งมีรายละเอียดขั้นตอนดังนี้

1. เลือกจุดศูนย์กลางกลุ่มแรก  $c_1$  โดยการสุ่มจากชุดข้อมูลทั้งหมดด้วยความน่าจะเป็นที่เท่ากัน
2. สุ่มเลือกจุดศูนย์กลางกลุ่มกลุ่มถัดไป  $c_j$  จากข้อมูลที่เหลือด้วยค่าความน่าจะเป็นที่มีค่ามากที่สุด สำหรับข้อมูลแต่ละ  $X_i$  ค่าความน่าจะเป็นเป็นดังสมการที่ (2)

$$P(c_j) = \frac{dist(X_i, C)^2}{\sum_{m=1}^n dist(x_m, C)^2} \quad (2)$$

โดยที่

$C$  คือ เซตของจุดศูนย์กลางกลุ่ม  $C_j$

$dist(X_i, C)$  คือ ระยะห่างระหว่างข้อมูล  $X_i$  กับจุดศูนย์กลางกลุ่มที่ใกล้ที่สุด

3. ทำซ้ำในขั้นตอนที่ 2. จนกระทั่งได้จุดศูนย์กลางกลุ่มครบ  $k$  ตัว

4. เมื่อได้จุดศูนย์กลางเริ่มต้นครบตามจำนวนกลุ่มที่กำหนดจากนั้นดำเนินการแบ่งกลุ่มเหมือนขั้นตอนการแบ่งกลุ่มแบบเคมีนส์

### 2.3 ขั้นตอนการแบ่งกลุ่มแบบเคมีดอยด์ (K-medoids)

Hae-Sang และ Chi-Hyuck [7] ได้นำเสนอวิธีการแบ่งกลุ่มแบบเคมีดอยด์ ซึ่งขั้นตอนการแบ่งกลุ่มนี้คล้ายกับการแบ่งกลุ่มแบบเคมีนส์ จะแตกต่างกันตรงขั้นตอนการปรับปรุงจุดศูนย์กลางกลุ่ม โดยจะคัดเลือกจากข้อมูลที่เหมาะสมที่จะเป็นจุดกึ่งกลางของกลุ่มมากที่สุดเป็นจุดศูนย์กลางกลุ่มใหม่ แทนการใช้ค่าเฉลี่ยของกลุ่มแบบของวิธีเคมีนส์ ทำให้วิธีนี้ทนทานต่อข้อมูลรบกวน (Noise) หรือ ข้อมูลที่ผิดปกติ (Outlier) มากกว่าวิธีการแบ่งกลุ่มแบบเคมีนส์ ซึ่งมีขั้นตอนดังนี้

1. การเลือกจุดศูนย์กลางกลุ่มเริ่มต้นที่เรียกว่า Medoids

1.1) คำนวณระยะห่างระหว่างข้อมูลทั้งหมดทุกคู่ด้วย  
มาตรวัดระยะแบบยุคลิด ดังสมการที่ (1)

1.2) สำหรับเซตข้อมูล  $X$  คำนวณค่า  $P_{X_j}$  ดังสมการที่ (3)

$$P_{X_j} = \frac{\sum_{i=1}^n \text{dist}(X_i, X_j)}{\sum_{i=1}^n \text{dist}(X_i, X_j)}, j = 1, \dots, n \quad (3)$$

โดยที่

$n$  คือ จำนวนข้อมูลทั้งหมด

$\text{dist}(X_i, X_j)$  คือ ระยะห่างระหว่างข้อมูลที่  $i$  กับ  $j$

$\text{dist}(X_i, X_l)$  คือ ระยะห่างระหว่างข้อมูลที่  $i$  กับ  $l$

1.3) เรียงลำดับค่า  $P_{X_j}$  จากค่าน้อยไปหาค่ามาก แล้ว  
เลือกข้อมูลที่มีค่า  $P_{X_j}$  ที่มีค่าน้อยสุดจำนวน  $k$  ตัวแรก เพื่อ  
เป็นจุดศูนย์กลางกลุ่ม Medoids เริ่มต้นสำหรับการแบ่งกลุ่ม

1.4) กำหนดให้ข้อมูลแต่ละตัวอยู่ในกลุ่มที่ใกล้กับ  
จุดศูนย์กลางกลุ่ม Medoids นั้นๆ มากที่สุดด้วย  
ตัววัดระยะห่างยุคลิด

1.5) คำนวณผลรวมของระยะห่างจากข้อมูลทุกตัวกับ  
จุดศูนย์กลางกลุ่ม Medoids ของข้อมูลนั้นๆ ทุกตัว

2. ปรับปรุงจุดศูนย์กลางกลุ่ม Medoids โดยเลือกจาก  
ข้อมูลที่มีผลรวมของระยะห่างกับข้อมูลอื่นๆ ในกลุ่ม  
ข้อมูลเดียวกันน้อยที่สุด

3. การกำหนดกลุ่มใหม่ให้แต่ละข้อมูล

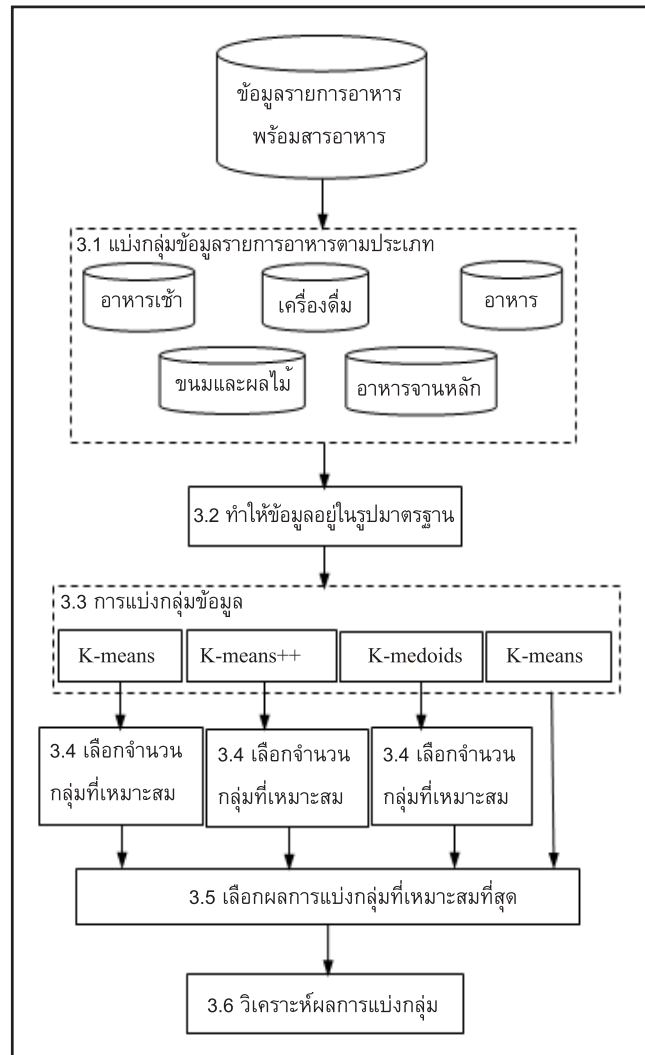
3.1) กำหนดให้ข้อมูลแต่ละตัวอยู่ในกลุ่มที่ใกล้กับ  
จุดศูนย์กลางกลุ่ม Medoids มากที่สุดด้วยตัววัดระยะห่าง  
ยุคลิด

3.2) คำนวณผลรวมของระยะห่างจากข้อมูลทุกตัวกับ  
จุดศูนย์กลางกลุ่ม Medoids ของข้อมูลนั้นๆ ถ้าผลรวมมีค่า  
เท่ากับผลรวมจากรอบที่แล้วจะหยุดการทำงาน แต่ถ้าไม่ให้  
กลับไปทำซ้ำในขั้นตอนข้อที่ 2

### 3. วิธีการดำเนินการวิจัย

งานวิจัยนี้เป็นการทดลองเพื่อหาวิธีแบ่งกลุ่มรายการ  
อาหารที่เหมาะสมที่ใช้สำหรับการวิเคราะห์รายการอาหาร  
โดยได้ทดลองใช้วิธีการแบ่งกลุ่มแบบเคมีนส์และวิธีการ  
แบ่งกลุ่มที่มีการปรับปรุงประสิทธิภาพของการแบ่งกลุ่มแบบ  
เคมีนส์ ได้แก่ การแบ่งกลุ่มแบบเคมีนส์พลัสพลัส การแบ่ง  
กลุ่มแบบเคมีนส์ดอยด์ และการแบ่งกลุ่มแบบเอ็กซ์มิน รวม

ทั้งหมด 4 วิธีการด้วยกัน โดยมีภาพรวมในการทำงาน  
ดังภาพที่ 1



ภาพที่ 1 ขั้นตอนวิธีการดำเนินงาน

#### 3.1 แบ่งกลุ่มข้อมูลรายการอาหารตามประเภท

ฐานข้อมูลรายการอาหารและสารอาหารที่ใช้ในงานวิจัย  
นี้ได้มาจากโปรแกรม Inmucal-Nutritions ของสถาบัน  
โภชนาการ มหาวิทยาลัยมหิดล [9] และสารอาหารที่ใช้เป็น  
ปัจจัยสำหรับแบ่งกลุ่มเป็นสารอาหารที่ถูกพิจารณาโดย  
นักโภชนาการแล้วว่า มีผลต่อโรคเบาหวาน ซึ่งอ้างอิงงานวิจัย  
[3] ได้แก่ คาร์โบไฮเดรต (กรัม) พลังงาน (กิโลแคลอรี)  
ใยอาหาร (กรัม) ไขมัน (กรัม) โปรตีน (กรัม) วิตามินอี  
(ไมโครกรัม) วิตามินซี (มิลลิกรัม) และไทอามีนหรือ  
วิตามินบีหนึ่ง (มิลลิกรัม) ตามลำดับ ซึ่งข้อมูลรายการอาหาร  
จะแบ่งออกเป็น 5 ประเภท คือ 1) อาหารเช้า จำนวน 1049  
รายการ 2) ขนมหวานและผลไม้ จำนวน 263 รายการ





3) เครื่องดื่ม จำนวน 105 รายการ 4) อาหารจานหลัก จำนวน 232 รายการ 5) อาหารว่าง จำนวน 92 รายการ รวมทั้งหมด 1,741 รายการ ดังตัวอย่างในตารางที่ 1 ที่ประกอบด้วยรหัสรายการอาหารและสารอาหาร 8 ชนิด

ตารางที่ 1 ตัวอย่างข้อมูลรายการอาหาร

| รหัส   | พลังงาน | คาร์โบไฮเดรต | ไขมัน | โปรตีน | โทอะมีน | วิตามินซี | วิตามินอี | ใยอาหาร |
|--------|---------|--------------|-------|--------|---------|-----------|-----------|---------|
| 900001 | 212.8   | 19.12        | 8.96  | 11.28  | 0.16    | 0.12      | 0.1       | 0       |
| 900002 | 244.8   | 28.32        | 8.96  | 10.48  | 0.14    | 0.12      | 0         | 0       |
| 900003 | 256.8   | 32.52        | 8.56  | 10.28  | 0.14    | 1.44      | 0.08      | 0       |
| 900004 | 176.2   | 17.2         | 7.18  | 8.64   | 0.122   | 0.12      | 0.07      | 0       |
| 900005 | 224.8   | 21.12        | 9.16  | 11.48  | 0.16    | 0.12      | 0         | 0       |
| 900006 | 244.8   | 28.72        | 9.16  | 10.28  | 0.12    | 0.12      | 0         | 0       |
| 900007 | 248.8   | 32.52        | 8.96  | 9.08   | 0.12    | 0.12      | 0         | 0       |
| 900008 | 213.3   | 25.07        | 11.81 | 7.13   | 0.205   | 2.145     | 0         | 0       |
| 900009 | 293.3   | 35.57        | 10.06 | 12.38  | 0.155   | 0.12      | 0         | 0       |

### 3.2 การทำข้อมูลให้เป็นมาตรฐาน

เนื่องจากข้อมูลรายการอาหารแต่ละรายการมีค่าของสารอาหารที่แตกต่างกันมาก ดังตารางที่ 2 ดังนั้นจึงจำเป็นต้องปรับขอบเขตของข้อมูลรายการอาหารให้อยู่ในช่วงที่เหมาะสมต่อการนำไปแบ่งกลุ่ม โดยให้รายการอาหารในแต่ละประเภทปรับค่าของสารอาหารให้อยู่ในขอบเขตในช่วง [0, 1] วิธีการทำข้อมูลให้เป็นบรรทัดฐานนั้นมีหลายวิธี แต่วิธีที่นิยมนำมาใช้กันอย่างแพร่หลายคือ การแปลงค่าข้อมูลในลักษณะเป็นเชิงเส้นที่เรียกว่า Min–Max Normalization [10] โดยคำนวณจากสมการที่ (4) ตารางที่ 3 เป็นตัวอย่างของข้อมูลหลังการทำให้อยู่ในรูปมาตรฐานแล้ว

ตารางที่ 2 ค่าที่มากที่สุดและน้อยที่สุดของสารอาหารแต่ละชนิดของประเภทอาหารเข้า

|           | พลังงาน | คาร์โบไฮเดรต | ไขมัน | โปรตีน | โทอะมีน | วิตามินซี | วิตามินอี | ใยอาหาร |
|-----------|---------|--------------|-------|--------|---------|-----------|-----------|---------|
| $X_{max}$ | 293.3   | 35.57        | 11.81 | 12.38  | 0.205   | 2.145     | 0.1       | 0       |
| $X_{min}$ | 176.2   | 17.2         | 7.18  | 7.13   | 0.12    | 0.12      | 0         | 0       |

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

$X_{new}$  คือ รายการอาหารที่ผ่านการปรับให้อยู่ในรูปมาตรฐานแล้ว

$X$  คือ รายการอาหารที่ยังไม่ผ่านการปรับให้อยู่ในรูปมาตรฐาน

$X_{min}$  คือ ค่าของสารอาหารที่น้อยที่สุด

$X_{max}$  คือ ค่าของสารอาหารที่มากที่สุด

ตารางที่ 3 ตัวอย่างข้อมูลหลังการทำให้อยู่ในรูปมาตรฐาน

| รหัส   | พลังงาน | คาร์โบไฮเดรต | ไขมัน | โปรตีน | โทอะมีน | วิตามินซี | วิตามินอี | ใยอาหาร |
|--------|---------|--------------|-------|--------|---------|-----------|-----------|---------|
| 900001 | 0.31255 | 0.104518236  | 0.384 | 0.7905 | 0.47059 | 0         | 1         | 0       |
| 900002 | 0.58582 | 0.605334785  | 0.384 | 0.6381 | 0.23529 | 0         | 0         | 0       |
| 900003 | 0.6883  | 0.833968427  | 0.298 | 0.6    | 0.23529 | 0.65185   | 0.8       | 0       |
| 900004 | 0       | 0            | 0     | 0.2876 | 0.02353 | 0         | 0.7       | 0       |
| 900005 | 0.41503 | 0.213391399  | 0.428 | 0.8286 | 0.47059 | 0         | 0         | 0       |
| 900006 | 0.58582 | 0.627109418  | 0.428 | 0.6    | 0       | 0         | 0         | 0       |
| 900007 | 0.61998 | 0.833968427  | 0.384 | 0.3714 | 0       | 0         | 0         | 0       |
| 900008 | 0.31682 | 0.428415895  | 1     | 0      | 1       | 1         | 0         | 0       |
| 900009 | 1       | 1            | 0.622 | 1      | 0.41176 | 0         | 0         | 0       |

### 3.3 การแบ่งกลุ่มข้อมูล

ในงานวิจัยนี้ได้ทดลองการแบ่งกลุ่มข้อมูลโดยใช้เครื่องมือ Toolbox จากโปรแกรม MATLAB (R2015a) และโปรแกรม WEKA 3.6.9 การแบ่งกลุ่มข้อมูลทั้ง 4 วิธีจะกำหนดจำนวนกลุ่ม (k) ตั้งแต่ k = 4 จนถึง k = 20 และเนื่องจากวิธีการแบ่งกลุ่มแบบเคมีนส์และเคมีนส์พลัสพลัส เลือกจุดศูนย์กลางกลุ่มเริ่มต้นด้วยวิธีการสุ่มด้วยความน่าจะเป็นที่เท่ากัน จึงได้ทดลองการแบ่งกลุ่มซ้ำ 10 ครั้ง (10 runs) ในแต่ละ k แล้วเลือกผลการแบ่งกลุ่มในครั้งที่ให้ค่าเฉลี่ย Silhouette [11] มากที่สุด

### 3.4 เลือกจำนวนกลุ่มที่เหมาะสม

Silhouette [11] เป็นวิธีการวัดความหนาแน่นและการกระจายตัวของกลุ่มข้อมูล โดยวัดความใกล้เคียงกันของข้อมูลแต่ละตัวภายในกลุ่มเดียวกันเทียบกับข้อมูลที่อยู่ต่างกลุ่มที่ใกล้ที่สุด (Neighboring Clusters) ซึ่งวิธีวัดนี้ให้ค่าตัวเลขอยู่ในช่วง [-1, 1] โดยหากมีค่าเข้าใกล้ 1 หมายความว่าข้อมูลถูกกำหนดให้อยู่ในกลุ่มที่เหมาะสม ถ้ามีค่าเท่ากับ 0 หมายความว่าไม่สามารถบอกได้อย่างชัดเจนว่าข้อมูลควรจะถูกกลุ่มใด และถ้ามีค่าเข้าใกล้ -1 หมายความว่าข้อมูลอาจถูกกำหนดให้อยู่ผิดกลุ่ม ดังสมการที่ (5)

$$S(X_i) = \frac{b(X_i) - a(X_i)}{\max\{a(X_i), b(X_i)\}} \quad (5)$$

โดยที่

$a(X_i)$  คือ ค่าเฉลี่ยของระยะห่างระหว่างรายการอาหาร  $X_i$  กับรายการอาหาร  $X_j$  ทุกรายการที่อยู่ในกลุ่มเดียวกันกับ  $X_i$

$b(X_i)$  คือ ค่าเฉลี่ยที่น้อยที่สุดของระยะห่างระหว่างรายการอาหาร  $X_i$  กับรายการอาหาร  $X_j$  ทุกรายการในกลุ่มข้อมูลอื่นๆ แต่ละกลุ่ม

$S(X_i)$  คือ ค่า Silhouette ในแต่ละรายการอาหาร  $X_i$

$$S(K) = \max_{k=1, \dots, 20} \frac{1}{n} \sum_{i=1}^n S(X_i) \quad (6)$$

โดยที่

$S(K)$  คือ ค่าเฉลี่ยของ Silhouette ที่มากที่สุดจากการกำหนดจำนวนกลุ่ม ( $k$ )

$n$  คือ จำนวนข้อมูลรายการอาหารทั้งหมด

### 3.5 เลือกผลการแบ่งกลุ่มที่เหมาะสมที่สุด

เมื่อได้จำนวนกลุ่มที่เหมาะสมกับรายการอาหารแต่ละประเภทแล้ว ขั้นตอนต่อมาคือ การคัดเลือกวิธีการแบ่งกลุ่มที่เหมาะสมที่สุดด้วยการวัดประสิทธิภาพการแบ่งกลุ่ม ได้แก่

Davies-Bouldin Index (DB) และ R-square Index (RS)

#### การวัดค่าแตกต่าง Davies-Bouldin Index (DB)

Davies-Bouldin Index (DB) [12] เป็นการวัดคุณภาพของการแบ่งกลุ่มโดยวัดความคล้ายกันของกลุ่มข้อมูล โดยคำนวณจากอัตราส่วนระหว่างผลรวมของการกระจายตัวของข้อมูลในกลุ่มและระยะห่างระหว่างกลุ่ม ซึ่งการแบ่งกลุ่มที่เหมาะสมข้อมูลภายในกลุ่มควรจะมีการกระจายคล้ายคลึงของข้อมูลภายในกลุ่มกันที่น้อย และมีค่าระยะห่างระหว่างจุดศูนย์กลางของระหว่างกลุ่มข้อมูลทีมาก ดังนั้นวิธีการแบ่งกลุ่มที่เหมาะสมที่สุดคือวิธีการแบ่งกลุ่มที่ให้ค่า DB ที่น้อยที่สุด

$$R_{c_i, c_j} = \frac{S_{c_i} + S_{c_j}}{\text{dist}_{c_i, c_j}} \quad (7)$$

โดยที่

$$S_{c_i} = \frac{1}{\|c_i\|} \sum_{X \in c_i} \text{dist}(X, c_i)$$

$$DB = \frac{1}{n} \sum_{i=1}^n R_i \quad (8)$$

โดยที่  $R_i = \max_{j=1, \dots, n_c, i \neq j} (R_{c_i, c_j})$   $i = 1, \dots, n_c$  และ

$R_{c_i, c_j}$  คือ ความคล้ายกันของกลุ่มที่  $i$  และกลุ่มที่  $j$

$\text{dist}_{c_i, c_j}$  คือ ระยะห่างระหว่างจุดศูนย์กลางกลุ่มที่  $i$  และกลุ่มที่  $j$

$c_i, c_j$  คือ จุดศูนย์กลางกลุ่มที่  $i$  และ  $j$

$S_{c_i}$  คือ ค่าระยะห่างเฉลี่ยของรายการอาหารที่เป็นสมาชิกกลุ่มที่  $i$  กับจุดศูนย์กลางกลุ่มที่  $i$

$S_{c_j}$  คือ ค่าระยะห่างเฉลี่ยของรายการอาหารที่เป็นสมาชิกกลุ่มที่  $j$  กับจุดศูนย์กลางกลุ่มที่  $j$

$\|c_i\|$  คือ จำนวนรายการอาหารในกลุ่มที่  $i$

$n_c$  คือ จำนวนกลุ่มข้อมูลทั้งหมด

#### การวัด R-square Index (RS)

R-square Index (RS) [12] เป็นการวัดค่าความแตกต่างระหว่างกลุ่ม ถ้าค่าระยะห่างระหว่างกลุ่มมีค่าที่มากแสดงว่ามีการแบ่งกลุ่มที่ค่อนข้างเหมาะสม โดยที่ค่า RS อยู่ในช่วง 0 ถึง 1 โดยที่ค่าเท่ากับ 0 หมายถึงไม่มีความแตกต่างระหว่างกลุ่มข้อมูล และมีค่าเท่ากับ 1 หมายถึงมีความแตกต่างระหว่างกลุ่มข้อมูลอย่างมีนัยสำคัญ

$$RS = \frac{SS_t - SS_w}{SS_t} \quad (9)$$

$$S_t = \sum_{j=1}^d \sum_{l=1}^{n_j} (x_l - \bar{x}_j)^2 \quad (10)$$

$$SS_w = \sum_{i=1 \dots n_c} \sum_{j=1 \dots d}^{n_j} (x_l - \bar{x}_j)^2 \quad (11)$$

โดยที่

$d$  คือ จำนวนสารอาหาร

$n_j$  คือ จำนวนสมาชิกในสารอาหารที่  $j$  จากเซตของรายการอาหารทั้งหมด

$\bar{x}_j$  คือ ค่าเฉลี่ยข้อมูลรายการอาหารในสารอาหารที่  $j$

$\bar{x}_{ij}$  คือ ค่าเฉลี่ยข้อมูลรายการอาหารในกลุ่มที่  $i$  สารอาหารที่  $j$

$n_{ij}$  คือ จำนวนสมาชิกในกลุ่มที่  $i$  และสารอาหาร  $j$

$n_c$  คือ จำนวนกลุ่มข้อมูลทั้งหมด

ตารางที่ 4 เป็นผลการวัดประสิทธิภาพการแบ่งกลุ่มทั้ง 4 วิธีในแต่ละประเภทอาหาร เพื่อเลือกวิธีการแบ่งกลุ่มที่เหมาะสมที่สุดในแต่ละประเภทอาหาร ในงานวิจัยนี้พิจารณาจากผลการวัดประสิทธิภาพการแบ่งกลุ่มที่ได้จากวิธีการ DB และ RS เป็นหลักก่อน โดยจะเลือกวิธีการแบ่งกลุ่มที่ให้ผลการวัด DB และ RS ที่ดีที่สุด และให้ผลการวัดจากทั้งสองวิธีที่ตรงกัน ดังจะเห็นได้จากแถวของอาหารประเภทขนมหวานและผลไม้ และอาหารว่าง แต่หากผลการวัดจากวิธี DB และ RS ไม่ตรงกัน ดังในแถวของประเภทอาหารเช้า



เครื่องดื่ม และอาหารจานหลัก ก็พิจารณาจากค่าเฉลี่ย Silhouette แทน ซึ่งผลการแบ่งกลุ่มที่เหมาะสมที่สุดในแต่ละประเภทแสดงได้ดังตารางที่ 5

**ตารางที่ 4** ผลการวัดประสิทธิภาพการแบ่งกลุ่มข้อมูลอาหาร 5 ประเภท

| ประเภทอาหาร     | วิธีการจัดกลุ่ม | จำนวนกลุ่ม (k) | ค่าเฉลี่ย Silhouette | DB          | RS          |
|-----------------|-----------------|----------------|----------------------|-------------|-------------|
| อาหารเช้า       | K-means         | 7              | 0.275861168          | 1.098537738 | 0.997410626 |
|                 | K-means++       | 8              | 0.44291732           | 1.068139869 | 0.996273411 |
|                 | K-medoids       | 11             | 0.4336               | 1.026114042 | 0.99603475  |
|                 | X-means         | 6              | 0.4201               | 1.069383294 | 0.997599603 |
| ขนมหวานและผลไม้ | K-means         | 4              | 0.343605075          | 1.235444    | 0.9848675   |
|                 | K-means++       | 5              | 0.488737836          | 1.2299181   | 0.9820329   |
|                 | K-medoids       | 4              | 0.453                | 1.24693     | 0.9779975   |
|                 | X-means         | 6              | 0.2613               | 1.1942297   | 0.9898138   |
| เครื่องดื่ม     | K-means         | 5              | 0.483649588          | 1.0030965   | 0.9475057   |
|                 | K-means++       | 5              | 0.697999524          | 0.8195694   | 0.9327795   |
|                 | K-medoids       | 7              | 0.6777               | 0.6588108   | 0.9327795   |
|                 | X-means         | 4              | 0.5952               | 1.2737332   | 0.9585647   |
| อาหารจานหลัก    | K-means         | 6              | 0.350542442          | 1.0098187   | 0.9907863   |
|                 | K-means++       | 4              | 0.523575324          | 1.1887116   | 0.9940363   |
|                 | K-medoids       | 5              | 0.4223               | 1.4855733   | 0.9914324   |
|                 | X-means         | 5              | 0.4296               | 1.3708156   | 0.9928768   |
| อาหารว่าง       | K-means         | 8              | 0.261220955          | 1.1931489   | 0.957109    |
|                 | K-means++       | 4              | 0.753390627          | 0.8623211   | 0.9685092   |
|                 | K-medoids       | 5              | 0.7426               | 1.1617006   | 0.9685092   |
|                 | X-means         | 4              | 0.51                 | 1.5177745   | 0.9626856   |

**ตารางที่ 5** ผลการวัดประสิทธิภาพการแบ่งกลุ่มข้อมูลและจำนวนกลุ่มที่เหมาะสมในประเภทอาหาร

| ประเภทรายการอาหาร | จำนวนกลุ่ม (k) | วิธีจัดการกลุ่ม |
|-------------------|----------------|-----------------|
| อาหารเช้า         | 11             | K-medoids       |
| ขนมและผลไม้       | 6              | X-means         |
| เครื่องดื่ม       | 7              | K-medoids       |
| อาหารจานหลัก      | 4              | K-means++       |
| ของว่าง           | 4              | K-means++       |

### 3.6 วิเคราะห์ผลการแบ่งกลุ่มข้อมูล

เมื่อได้จำนวนกลุ่มที่เหมาะสมในแต่ละประเภทอาหารแล้ว ขั้นตอนต่อมาเป็นขั้นตอนการวิเคราะห์ผลของการแบ่งกลุ่มว่าในแต่ละกลุ่มที่แบ่งออกมานั้น มีลักษณะเด่นของ

แต่ละสารอาหารเป็นอย่างไร โดยในงานวิจัยนี้จะแบ่งช่วงของสารอาหารออกเป็น 3 ระดับ ได้แก่ ระดับต่ำ ปานกลาง และสูง ที่คำนวณจาก (ค่าที่มากที่สุด-ค่าที่น้อยที่สุด)หารด้วย 3 ดังตัวอย่างในตารางที่ 6 การแบ่งช่วงสารอาหารของประเภทอาหารเช้า ส่วนประเภทอาหารอื่นๆ ก็พิจารณาในลักษณะทำนองเดียวกัน หลังจากนั้นข้อมูลสารอาหารที่เป็นจุดศูนย์กลางกลุ่มของแต่ละกลุ่มจะถูกเปลี่ยนระดับของสารอาหารตามช่วงที่กำหนดไว้ของแต่ละสารอาหาร

**ตารางที่ 6** การแบ่งช่วงของสารอาหารของประเภทอาหารเช้า

| สารอาหาร     | ต่ำ                 | ปานกลาง             | สูง                 |
|--------------|---------------------|---------------------|---------------------|
| พลังงาน      | 168.3854 - 303.3236 | 303.3237 - 438.2618 | 438.2619 - 573.2000 |
| คาร์โบไฮเดรต | 14.2500 - 31.0782   | 31.0783 - 47.9063   | 47.9064 - 64.7345   |
| ไขมัน        | 1.6824 - 12.3046    | 12.3047 - 22.9268   | 22.9269 - 33.5490   |
| โปรตีน       | 2.0600 - 11.0393    | 11.0394 - 20.0187   | 20.0188 - 28.9980   |
| โทเคมิน      | 0.0000 - 1.6414     | 1.6415 - 3.2828     | 3.2829 - 4.9242     |
| วิตามินซี    | 0.0000 - 15.7276    | 15.7277 - 31.4552   | 31.4553 - 47.1828   |
| วิตามินอี    | 0.0000 - 7.5702     | 7.5703 - 15.1404    | 15.1405 - 22.7106   |
| ใยอาหาร      | 0.0000 - 5.9057     | 5.9058 - 11.8113    | 11.8114 - 17.7170   |

**ตารางที่ 7** ลักษณะข้อมูลสารอาหารของแต่ละกลุ่มของประเภทอาหารเช้า

| กลุ่ม       | พลังงาน | คาร์โบไฮเดรต | ไขมัน   | โปรตีน  | โทเคมิน | วิตามินซี | วิตามินอี | ใยอาหาร |
|-------------|---------|--------------|---------|---------|---------|-----------|-----------|---------|
| กลุ่มที่ 1  | ปานกลาง | ปานกลาง      | ปานกลาง | ปานกลาง | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 2  | ปานกลาง | สูง          | ปานกลาง | ต่ำ     | สูง     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 3  | ต่ำ     | ปานกลาง      | ต่ำ     | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | สูง     |
| กลุ่มที่ 4  | ต่ำ     | ต่ำ          | ต่ำ     | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 5  | สูง     | ปานกลาง      | สูง     | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 6  | ปานกลาง | ปานกลาง      | ปานกลาง | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 7  | ต่ำ     | ปานกลาง      | ต่ำ     | ต่ำ     | ต่ำ     | ต่ำ       | สูง       | ต่ำ     |
| กลุ่มที่ 8  | ปานกลาง | ต่ำ          | ปานกลาง | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 9  | ต่ำ     | ปานกลาง      | ต่ำ     | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 10 | ปานกลาง | ปานกลาง      | ปานกลาง | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 11 | ต่ำ     | ต่ำ          | ปานกลาง | ต่ำ     | ต่ำ     | ปานกลาง   | ต่ำ       | ต่ำ     |

**ตารางที่ 8** ลักษณะข้อมูลสารอาหารของข้อมูลกลุ่มเครื่องดื่ม

| กลุ่ม      | พลังงาน | คาร์โบไฮเดรต | ไขมัน | โปรตีน  | โทเคมิน | วิตามินซี | วิตามินอี | ใยอาหาร |
|------------|---------|--------------|-------|---------|---------|-----------|-----------|---------|
| กลุ่มที่ 1 | ต่ำ     | ต่ำ          | ต่ำ   | ต่ำ     | ต่ำ     | ต่ำ       | สูง       | ต่ำ     |
| กลุ่มที่ 2 | ต่ำ     | ต่ำ          | ต่ำ   | ปานกลาง | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 3 | สูง     | สูง          | ต่ำ   | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 4 | ต่ำ     | ปานกลาง      | ต่ำ   | ต่ำ     | ต่ำ     | สูง       | ต่ำ       | สูง     |
| กลุ่มที่ 5 | สูง     | สูง          | สูง   | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | สูง     |
| กลุ่มที่ 6 | ต่ำ     | ต่ำ          | ต่ำ   | ต่ำ     | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 7 | ต่ำ     | ต่ำ          | ต่ำ   | ต่ำ     | ปานกลาง | ต่ำ       | ต่ำ       | ต่ำ     |

**ตารางที่ 9** ลักษณะข้อมูลสารอาหารของข้อมูลกลุ่มของอาหารว่าง

| กลุ่ม      | พลังงาน | คาร์โบไฮเดรต | ไขมัน   | โปรตีน | โทเคมิน | วิตามินซี | วิตามินอี | ใยอาหาร |
|------------|---------|--------------|---------|--------|---------|-----------|-----------|---------|
| กลุ่มที่ 1 | ต่ำ     | ต่ำ          | ต่ำ     | ต่ำ    | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 2 | ปานกลาง | ต่ำ          | ปานกลาง | ต่ำ    | ต่ำ     | ต่ำ       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 3 | ต่ำ     | ต่ำ          | ปานกลาง | ต่ำ    | ต่ำ     | สูง       | ต่ำ       | ต่ำ     |
| กลุ่มที่ 4 | ปานกลาง | ต่ำ          | สูง     | สูง    | ต่ำ     | ต่ำ       | สูง       | สูง     |

ตารางที่ 7 ถึงตารางที่ 9 เป็นตัวอย่างผลจากการเปลี่ยนศูนย์กลางกลุ่มของแต่ละกลุ่มไปเป็นระดับของสารอาหารตามช่วงที่กำหนดไว้ของแต่ละสารอาหารที่เป็นประเภทอาหารเข้า เครื่องดื่ม และอาหารว่างตามลำดับ ส่วนอาหารประเภทอื่นที่ไม่ได้แสดงก็พิจารณาในลักษณะทำนองเดียวกันนี้

#### 4. การวิเคราะห์คุณลักษณะเด่น

ในการวิเคราะห์กลุ่มอาหารของรายการอาหารทั้ง 5 ประเภทได้อ้างอิงข้อมูลด้านโภชนาการจากกองโภชนาการกรมอนามัย ซึ่งได้ผลการวิเคราะห์กลุ่มข้อมูลดังนี้

ประเภทอาหารเข้า จำนวน 11 กลุ่ม

กลุ่มที่ 1, 2, 3, 4, 6, 7, 8, 9, 10, 11 : เป็นกลุ่มอาหารที่สามารถรับประทานได้ แต่ต้องควบคุมปริมาณการรับประทาน เพราะระดับสารอาหารส่วนใหญ่อยู่ในระดับปานกลางถึงต่ำในกลุ่มที่ 2 และ 3 แม้จะมีระดับคาร์โบไฮเดรตปานกลางและสูง แต่ก็ยังมีปริมาณไขมันและใยอาหารที่สูงด้วย จึงยังคงแนะนำให้รับประทาน เพราะใยอาหารมีส่วนช่วยลดการดูดซึมของน้ำตาลในเลือดได้ และไขมันยังช่วยย่อยอาหารจำพวกคาร์โบไฮเดรตได้ดี

กลุ่มที่ 5 : เป็นกลุ่มที่ควรหลีกเลี่ยง หรือรับประทานได้เพียงเล็กน้อยเท่านั้น เนื่องจากเป็นกลุ่มอาหารที่มีปริมาณพลังงานและไขมันสูง อาจส่งผลกระทบต่อโรคเบาหวานให้มีความรุนแรงยิ่งขึ้น

ประเภทขนมหวานและผลไม้ จำนวน 6 กลุ่ม

กลุ่มที่ 1, 2 : เป็นกลุ่มอาหารสามารถรับประทานได้ แต่ควรรับประทานในปริมาณที่จำกัด เนื่องจากมีปริมาณพลังงาน คาร์โบไฮเดรต และไขมัน ในระดับปานกลาง และยังมีปริมาณวิตามินซีสูง ซึ่งช่วยในลดความต้องการอินซูลินได้

กลุ่มที่ 3, 4, 5, 6 : เป็นกลุ่มอาหารที่สามารถรับประทานได้ปกติ เนื่องจากระดับสารอาหารส่วนใหญ่อยู่ในระดับต่ำ

ประเภทเครื่องดื่ม จำนวน 7 กลุ่ม

กลุ่มที่ 1, 6, 7 : เป็นกลุ่มเครื่องดื่มที่สามารถบริโภคได้ เพราะระดับสารอาหารต่ำ ส่วนใหญ่อยู่ในระดับต่ำ

กลุ่มที่ 2, 4 : เป็นกลุ่มที่บริโภคได้แต่ควรมีการควบคุมปริมาณเนื่องจากยังคงมีระดับคาร์โบไฮเดรตในระดับปานกลาง

กลุ่มที่ 3, 5 : เป็นกลุ่มเครื่องดื่มที่ควรหลีกเลี่ยง หรือบริโภคเพียงเล็กน้อย เนื่องจากมีปริมาณพลังงาน คาร์โบไฮเดรต และไขมันสูง อีกทั้งเครื่องดื่มยังสามารถดูดซึมเข้าสู่ร่างกายได้รวดเร็ว จึงยังมีผลทำให้ระดับน้ำตาลในเลือดมีการเปลี่ยนแปลงอย่างรวดเร็วเพิ่มขึ้น

ประเภทอาหารจานหลัก จำนวน 4 กลุ่ม

กลุ่มที่ 1, 2 : เป็นกลุ่มที่แนะนำให้รับประทานได้ เนื่องจากมีระดับของวิตามินซีและวิตามินอีสูง ซึ่งช่วยในเรื่องขยายเส้นเลือดและการจับตัวเป็นลิ่มเลือด จึงทำให้เลือดไหลเวียนได้ดี

กลุ่มที่ 3, 4 : เป็นกลุ่มที่แนะนำให้รับประทานได้ แต่ควรควบคุมปริมาณในการรับประทาน

ประเภทอาหารว่าง จำนวน 4 กลุ่ม

กลุ่มที่ 1, 2, 3, 4 : เป็นกลุ่มที่สามารถรับประทานได้ แต่ยังคงต้องควบคุมปริมาณการรับประทาน ในกลุ่มที่ 4 แม้ว่าจะมีปริมาณไขมันที่สูง แต่เนื่องจากมีปริมาณของโปรตีน วิตามินอีและใยอาหารที่สูงจึงยังคงแนะนำให้รับประทาน

#### 5. สรุป

การควบคุมการรับประทานอาหารเป็นปัจจัยสำคัญที่จะช่วยรักษาและลดความรุนแรงของโรคเบาหวาน ดังนั้นสิ่งที่สำคัญคือผู้ป่วยจำเป็นต้องทราบว่าอาหารที่จะรับประทานนั้นจัดอยู่ประเภทใด งานวิจัยนี้จึงมีแนวคิดในการใช้วิธีแบ่งกลุ่มข้อมูลเข้ามาประยุกต์ใช้ โดยการแบ่งข้อมูลรายการอาหารในแต่ละประเภทออกเป็นกลุ่มๆ ตามสารอาหารที่มีความคล้ายคลึงกัน โดยในงานวิจัยนี้ได้ทดลองวิธีการแบ่งกลุ่มข้อมูลด้วยกัน 4 วิธี ได้แก่ การแบ่งกลุ่มแบบเคมีนส์แบบเคมีนส์พลัสพลัส แบบเคมีนส์ดอยด์ และแบบเอ็กซ์มิน จากนั้นใช้ค่าเฉลี่ย Silhouette ตัววัดประสิทธิภาพการแบ่งกลุ่มด้วยวิธี DB และ RS หาจำนวนกลุ่มและวิธีการแบ่งกลุ่มที่เหมาะสมกับข้อมูลอาหารในแต่ละประเภท ซึ่งผลการแบ่งกลุ่มได้ผลดังนี้ประเภทอาหารเข้าแบ่งได้ 11 กลุ่ม ประเภทเครื่องดื่มแบ่งได้ 7 กลุ่ม จากวิธีการแบ่งกลุ่มแบบเคมีนส์ดอยด์ ประเภทขนมและผลไม้แบ่งได้ 6 กลุ่ม จากวิธีการแบ่งกลุ่มแบบเอ็กซ์มินส์ ประเภทอาหารจานหลักแบ่งได้ 4 กลุ่ม และประเภทอาหารว่างแบ่งได้ 4 กลุ่ม จากวิธีการแบ่งกลุ่มแบบเคมีนส์พลัสพลัส ผลการวิเคราะห์กลุ่มอาหารสามารถแบ่งกลุ่มอาหารได้ดังนี้ คือ กลุ่มที่ควรหลีกเลี่ยงหรือ





รับประทานในปริมาณที่น้อยๆ ได้แก่ ประเภทอาหารเข้ากลุ่มที่ 5 ประเภทเครื่องดื่มกลุ่มที่ 3 และ 5 ส่วนกลุ่มสามารถรับประทานได้ แต่ต้องมีการควบคุมปริมาณการรับประทาน ได้แก่ ประเภทอาหารเข้า กลุ่มที่ 1, 2, 3, 4, 6, 7, 8, 9, 10, 11 ประเภทขนมหวานและผลไม้ กลุ่มที่ 1, 2 ประเภทเครื่องดื่มกลุ่มที่ 2 และ 4 ประเภทอาหารจานหลัก กลุ่มที่ 3 และ 4 และอาหารในกลุ่มที่ผู้ป่วยเบาหวานสามารถรับประทานได้ปกติ ได้แก่ ประเภทขนมหวานและผลไม้ กลุ่มที่ 3, 4, 5 และ 6 ประเภทเครื่องดื่มกลุ่มที่ 1 และ 6

## 6. เอกสารอ้างอิง

- [1] W. Kriengsinyos. "Nutrition and Diabetes" 3rd edition, Bangkok, Sarakadee Magazine, 2008.
- [2] H. C. Li and W. M. Ko. "Automated food ontology construction mechanism for diabetes diet care." In *Proceedings of 2007 International Conference on Machine Learning and Cybernetics*, Hong Kong, pp. 2953-2958, 2007.
- [3] P. Maiyaporn, P. Pholkul, and S. Phimoltares. "Food recommendation system using clustering analysis for diabetic patients." In *Proceedings of 2010 International Conference on Information Science and Applications (ICISA)*, pp. 1-8, Seoul, Korea, 2010.
- [4] Y. Usthasopha, S. Phimoltares, and N. Cooharajanane. "Nutrition Counseling System and Food Menu Planning." In *Proceeding of 2010 International Conference on Information Science and Applications (ICISA)*, pp. 1-8, Seoul, Korea, 2010.
- [5] J. MacQueen. "Some methods for classification and analysis of multivariate observations." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1: Statistics, pp. 281-297, 1967.
- [6] D. Arthur and V. Sergei. "K-means++: The advantages of careful seeding." In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms*, pp. 1027-1035, 2007.
- [7] H.S. Park and C.H. Jun. "A simple and fast algorithm for K-medoids clustering." *Expert Systems with Applications*, Vol. 36, No. 2, pp. 3336-3341, 2009.
- [8] D. Pelleg and A.W. Moore. "X-means: Extending K-means with Efficient Estimation of the Number of Clusters." In *Proceedings of the 17th International Conference on Machine Learning*, pp. 727-734, 2000.
- [9] Institute of Nutrition, Mahidol University. *INMUCAL-Nutrients V3 Program*. Nakhon Pathom, Thailand, 2013.
- [10] P. N. Tan, M. Steinbach, and V. Kumar. "Introduction to data mining." Vol. 1. Boston: Pearson Addison Wesley, 2006.
- [11] P. J. Rousseeuw. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of computational and applied mathematics*, Vol. 20, pp.53-65, 1987.
- [12] F. Kovács, C. Legány, and A. Babos. "Cluster validity measurement techniques." In *Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence*, pp. 388-393, Budapest. 2005.