



**ระบบแนะนำสถานที่โดยการประยุกต์ใช้  
หลักการวิเคราะห์ข้อมูลโพรไฟล์ประชากรจากเครือข่ายสังคม  
Place Recommendation System Using  
Demographic Data Analysis Principle from Social Network**

ชรีญา แยมอคุลย์<sup>1</sup> นครทิพย์ พร้อมพูล<sup>1,2\*</sup> และ อรรถสิทธิ์ สุรฤกษ์<sup>1,3\*</sup>

**บทคัดย่อ**

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการสร้างกฎและเครื่องมือสำหรับการแนะนำสถานที่จากการจำแนกข้อมูลโพรไฟล์ประชากร ด้วยหลักวิธีการจำแนกประเภทเบสอย่างง่าย โดยใช้ข้อมูลโพรไฟล์ประชากรจากเครือข่ายสังคมเฟซบุ๊ก ทั้งหมด 7 รายการ ได้แก่ เพศ อายุ ความสัมพันธ์ การศึกษา สถานะการทำงาน ภูมิลำเนา และที่อยู่ปัจจุบัน และได้จำแนกสถานที่เพื่อการแนะนำเป็น 4 หมวดหมู่จากรายการหมวดหมู่ที่ได้ระบุไว้ในเฟซบุ๊ก ได้แก่ ร้านอาหาร โรงแรม สถานที่ท่องเที่ยวกลางแจ้ง และศูนย์การค้าและค้าปลีก พร้อมทั้งได้กำหนดรายการหมวดหมู่ย่อยของแต่ละหมวดหมู่ด้วย ในการระบุกลุ่มให้กับกลุ่มประชากรตัวอย่างเพื่อใช้เป็นข้อมูลสอนนั้นพิจารณาจากความถี่ของการไปเยือนสถานที่จำนวน 35,084 รายการ ของกลุ่มผู้ใช้ตัวอย่างจำนวน 600 รายที่ได้จากการรวบรวมข้อมูลจากเฟซบุ๊ก

การวัดประสิทธิผลจากค่าความแม่นยำของกฎที่สร้างขึ้นของแต่ละหมวดหมู่ของสถานที่ทั้ง 4 รายการ ได้แก่ ร้านอาหาร โรงแรม สถานที่ท่องเที่ยวกลางแจ้ง และศูนย์การค้าและค้าปลีก พบว่ามีค่าความแม่นยำเป็น 78.7 72.5 67.4 และ 76.7% ตามลำดับ เครื่องมือที่พัฒนาขึ้นสามารถสนับสนุนการสืบค้นข้อมูลสถานที่ในแต่ละหมวดหมู่จากโพรไฟล์ของผู้ใช้ได้ อีกทั้งสามารถสนับสนุนการปรับปรุงข้อมูลการเยือนสถานที่ของกลุ่มผู้ใช้ตัวอย่างเป้าหมายเพื่อนำมาใช้ในการปรับปรุงชุดข้อมูลสอน ให้ได้ระบบการแนะนำที่มีความทันสมัยของข้อมูลและสอดคล้องกับความสนใจในปัจจุบันของผู้ใช้

<sup>1</sup>ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ถนนพญาไท แขวงวังใหม่ เขตปทุมวัน กรุงเทพฯ 10330

\*Corresponding Author, E-mail: <sup>2</sup>Nakornthip.S@chula.ac.th, <sup>3</sup>Athasit.S@chula.ac.th

## ABSTRACT

The objective of this master project is to present rule and tool for place recommendation system using population demographic information from Facebook using Naïve Bayes classification. The demographic information is composed of 7 items; gender, age, relationship, education, occupation status, hometown and current address. The recommended places were classified into 4 types: restaurant, hotel, outdoor visiting place, and department store and retail store. Moreover, the subcategories of each category were defined. The frequency of place visiting based on user check-in information was used to identify each user group for training data. The total frequency of check-in transaction was 35,084 items from 600 Facebook's users.

The effectiveness of the proposed rule was used accuracy metric. The accuracy values of rule for recommended places; the restaurant, hotel, outdoor visiting place, and department and retail store; were 78.7, 72.5, 67.4 and 76.7% respectively. The developed tool based on the proposed rule could be used to search the recommended places using user profiles. In addition, this tool provided the updating function to adjust the check-in information of the target sample users in order to improving the training data set. This would enhance the recommendation system in term of the up-to-date information and user interest.

**คำสำคัญ:** ระบบแนะนำสถานที่ หลักวิธีการจำแนกประเภทเบย์อย่างง่าย

**Keywords:** Place recommendation system, Naïve Bayes classification

## บทนำ

เทคโนโลยีการสื่อสารในปัจจุบันมีการพัฒนาขึ้นอย่างรวดเร็ว การเข้าถึงอินเทอร์เน็ต (internet) สามารถทำได้สะดวกยิ่งขึ้น เนื่องจากมีอุปกรณ์ที่สนับสนุนการใช้งานอินเทอร์เน็ตมากมาย ทำให้ปริมาณผู้ใช้งานอินเทอร์เน็ตเพิ่มขึ้นเรื่อย ๆ จึงเกิดเป็นเครือข่ายสังคม (social network) ที่เป็นการกำหนดโครงสร้างทางสังคมของบุคคลที่มีความเกี่ยวข้องกันทั้งทางตรงและทางอ้อม มีพื้นฐานจากความสนใจที่เหมือนกัน (Jaideep, 2008) ในปัจจุบันมีบริษัทผู้ผลิตโปรแกรมประยุกต์หลายรายได้ผลิตโปรแกรมประยุกต์ในลักษณะของเครือข่ายสังคม

ออกมาตอบสนองความต้องการของผู้ใช้ในหลากหลายรูปแบบ เช่น เฟซบุ๊ก (Facebook) เป็นเครือข่ายสังคมที่มีจุดประสงค์เพื่อแบ่งปันข้อมูล รูปภาพ วิดีโอ หรือสถานที่ที่ผู้ใช้ไปเยือน เป็นต้น เฟซบุ๊กติดอันดับหนึ่งของการจัดอันดับเว็บไซต์สื่อสังคมด้านความนิยม ( [www.experian.com/hitwise/online-trends-social-media.html](http://www.experian.com/hitwise/online-trends-social-media.html).) จากสถิติเดือนมิถุนายน ค.ศ. 2013 มีผู้ใช้งานจำนวน 1.15 พันล้านคนต่อเดือนและ 699 ล้านคนต่อวัน ( <https://newsroom.fb.com/Key-Facts>) จากการที่มีโปรแกรมประยุกต์มากมายในลักษณะของเครือข่ายสังคม การแบ่งปันข้อมูลสู่โลกอินเทอร์เน็ตจึงเป็นเรื่องที่ง่ายอย่างยิ่ง ทำให้ข้อมูลมีการ

เติบโตขึ้นอย่างรวดเร็ว ข้อมูลบนโลกอินเทอร์เน็ตจึงมีมากมายและมีแหล่งที่มาของข้อมูลที่หลากหลาย ดังนั้น การค้นหาข้อมูลได้อย่างรวดเร็วและตรงตามความต้องการจากแหล่งข้อมูลขนาดใหญ่ นั้น ทำให้เกิดปัญหาของผู้ใช้งาน

ในชีวิตประจำวันของเรามักพบกับ การตัดสินใจเลือกสถานที่ในสถานการณ์ต่างๆ ยกตัวอย่าง เช่น การเลือกร้านอาหารมื้อเที่ยง หรือการเลือกที่พักสำหรับวันหยุด เมื่อใช้บริการการค้นหาสถานที่จากอินเทอร์เน็ต ผู้ใช้จะพบรายการผลลัพธ์ที่ได้จากการค้นหานั้นมากมาย ทั้งนี้เพื่อเป็นการรวบรวมข้อมูลสถานที่ไว้ในที่เดียวและเพื่อเป็นการสนับสนุนการค้นหาที่มีความเหมาะสมกับผู้ใช้ ผู้วิจัยจึงต้องการนำเสนอระบบแนะนำ (recommender system) เพื่อเป็นการสนับสนุนการค้นหาสถานที่จากอินเทอร์เน็ตให้มีความสะดวกและรวดเร็วยิ่งขึ้น และได้ผลลัพธ์ที่ตรงกับความ ต้องการของผู้ใช้

ระบบแนะนำเป็นระบบที่ใช้แนะนำข้อมูลตามความต้องการของผู้ใช้ หลักวิธีการที่นำมาประยุกต์ใช้กับระบบแนะนำนั้นมีหลากหลายวิธี ในงานวิจัยนี้ เลือกใช้การแนะนำด้วยการใช้ข้อมูลประชากร (demographic) เช่น เพศ อายุ การศึกษา เพื่อนำมาประกอบการแนะนำ ซึ่งวิธีการวิเคราะห์ด้วยข้อมูลประชากรไม่จำเป็นต้องอาศัยข้อมูลอื่น เช่น ประวัติการเข้าถึงเว็บไซต์ของผู้ใช้ในอดีต รายการสถานที่ที่ผู้ใช้เคยแสดงความเห็นว่าถูกใจ และข้อมูลรายละเอียดของสิ่งที่ต้องการค้นหา เป็นต้น ทำให้ผู้ใช้ใหม่ที่ไม่เคยมีประวัติการใช้งานระบบสามารถใช้งานระบบแนะนำที่ใช้หลักของข้อมูลการวิเคราะห์ด้วยข้อมูลประชากรได้

ระบบแนะนำที่นำเสนอนี้เหมาะสำหรับผู้ใช้ที่ต้องการค้นหาสถานที่ 4 ประเภทได้แก่ ร้านอาหาร โรงแรม สถานที่ท่องเที่ยวกลางแจ้ง และศูนย์การค้า และคาเฟ่ โดยการนำข้อมูลโพรไฟล์ของผู้ใช้มา

วิเคราะห์ด้วยหลักวิธีการจำแนกประเภทเบสส์อย่างง่าย เพื่อให้ได้สถานที่ที่เหมาะสมกับผู้ใช้ และเพื่อความสะดวกในการค้นหาสถานที่ภายในเว็บเดียว

รายละเอียดที่จะกล่าวถึงในหัวข้อต่อไปจะกล่าวถึงทฤษฎีที่เกี่ยวข้อง งานวิจัยที่เกี่ยวข้อง การสร้างกฎสำหรับการจำแนกประเภทเบสส์อย่างง่าย การสร้างระบบแนะนำสถานที่ และสรุปผลงานวิจัยและแนวทางการดำเนินการในอนาคต

## ทฤษฎีที่เกี่ยวข้อง

### 1 หลักวิธีการจำแนกประเภทเบสส์อย่างง่าย

หลักวิธีการจำแนกประเภทเบสส์อย่างง่าย (บุญเสริม, 2548) มีรากฐานมาจากทฤษฎีบทของเบสส์ มีจุดประสงค์เพื่อสร้างรูปแบบสำหรับเลือกสมมติฐาน ที่ถูกต้องที่สุดโดยใช้หลักความน่าจะเป็นเข้ามา การจำแนกประเภทเบสส์อย่างง่าย สามารถคำนวณได้ด้วยสมการที่ (1)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

โดยกำหนดให้ A,B เป็นเหตุการณ์หนึ่ง ๆ P(A|B) คือความน่าจะเป็น A เมื่อรู้ B โดยมี เงื่อนไขว่า B ต้องเกิดขึ้นแล้ว ซึ่งเรียก P(A) ว่าเป็นความน่าจะเป็นก่อน (prior probability) เป็นค่าความน่าจะเป็น ที่ได้จากข้อมูลเบื้องต้น และเรียก P(A|B) ว่าเป็นความน่าจะเป็นหลัง (posterior probability) เป็นค่าความน่าจะเป็นก่อนที่ถูกปรับด้วยข้อมูลที่เพิ่มขึ้น ในกรณีที่นำทฤษฎีบทของเบสส์มาใช้กับแมชชีนเลิร์นนิง สิ่งที่น่าสนใจคือการหาความน่าจะเป็นของโอกาสที่จะเกิดสมมติฐานที่เราสนใจ จากข้อมูลชุดตัวอย่างหรือชุดข้อมูลสอน โดยมีการคำนวณตามสมการที่ (2)

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (2)$$

โดยกำหนดให้  $h$  เป็นข้อสมมติฐานและ  $D$  เป็นชุดข้อมูลสอน  $P(h)$  เป็นค่า ความน่าจะเป็นของสมมติฐาน  $h$  ที่ยังไม่ทราบข้อมูล  $D$  ส่วน  $P(D|h)$  เป็นค่าความน่าจะเป็นภายหลัง ซึ่งเป็นค่าความน่าจะเป็นของสมมติฐาน  $h$  จะเป็นจริงก็ต่อเมื่อ  $D$  เป็นจริง ส่วน  $P(D)$  เป็นค่าความน่าจะเป็นของชุดข้อมูลตัวอย่างหรือข้อมูลสอน  $D$  ในการหาค่าสมมติฐานที่ดีที่สุดเรียกว่า สมมติฐานภายหลังมากที่สุด-เอ็มเอพี (maximum A posterior hypothesis: MAP) ซึ่งมีการนิยามดังสมการที่ (3)

$$h_{\text{map}} = \underset{h \in H}{\operatorname{argmax}} P(h|D) \quad (3)$$

$$= \underset{h \in H}{\operatorname{argmax}} \frac{P(D|h)P(h)}{P(D)}$$

โดยที่  $H$  เป็นปริภูมิของสมมติฐานทั้งหมด  $\operatorname{argmax} f(x)$  เป็น ฟังก์ชันที่คืนค่า  $x$  ที่ทำให้  $f(x)$  สูงสุด เนื่องจากว่าสำหรับ  $h \in H$  ทุกตัว มีค่า  $P(D)$  เท่ากันหมดจึงสามารถลดรูปได้ดังสมการที่ (4)

$$h_{\text{map}} = \underset{h \in H}{\operatorname{argmax}} P(D|h)P(h) \quad (4)$$

หลักวิธีการจำแนกประเภทเบสอย่างง่ายเป็นหนึ่งในหลายวิธี ของแมชชีนเลิร์นนิงด้วยเทคนิคการเรียนรู้แบบมีผู้สอน โดยการหาความน่าจะเป็น ถ้าสมมติให้  $A_1, A_2, \dots, A_n$  เป็นคุณสมบัติของชุดข้อมูลตัวอย่าง หรือชุดข้อมูลสอน จะได้ว่าค่า (ประเภท) ที่น่าจะเป็นที่สุดของตัวอย่าง  $x$  ดัง สมการที่ (5)

$$V_{\text{map}} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j|a_1, a_2, \dots, a_n) \quad (5)$$

$$= \underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2, \dots, a_n|v_j)P(v_j)}{P(a_1, a_2, \dots, a_n)}$$

โดยที่  $a_i$  แทนค่าคุณสมบัติของ  $A_i$  ส่วน  $V$  เป็นเซตของประเภท หรือค่าที่เป็นไปได้ของ  $x$  และเนื่องจากการกำหนดให้คุณสมบัติของแต่ละตัวไม่ขึ้น

(เป็นอิสระ) กับคุณสมบัติอื่นทำให้เราแทนค่า  $P(v_j|a_1, a_2, \dots, a_n)$  ได้ดังสมการที่ (6)

$$P(a_1, a_2, \dots, a_n|v_j) = \prod_{i=1}^n P(a_i|v_j) \quad (6)$$

โดยที่  $\prod$  คือการนำค่า  $P(a_i|v_j)$  ทั้งหมดมาคูณกัน ดังนั้นความน่าจะเป็นทางด้านซ้ายของสมการจะมีค่าเท่ากับผลคูณความน่าจะเป็นทางขวาของสมการก็ต่อเมื่อคุณสมบัติของ  $a_1, a_2, \dots, a_n$  ต้องไม่ขึ้นต่อกัน เช่น สีมผมจะไม่ขึ้นกับความสูงของคน อายุจะไม่ขึ้นต่อเพศ เป็นต้น ดังนั้นการ จำแนกประเภทเบสอย่างง่ายคือสมการที่ (7)

$$V_{\text{NB}} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \times \prod_{i=1}^n P(a_i|v_j) \quad (7)$$

## 2 การเก็บรวบรวมข้อมูลจากเครือข่ายสังคมเพชบุรี

การนำข้อมูลออกจากเพชบุรีจำเป็นต้องใช้ ส่วนต่อประสานโปรแกรมประยุกต์ (application program interface: API) เพื่อติดต่อกับเพชบุรีที่เรียกว่า เพชบุรีโอเพนกราฟ (F. Müller and F. Thiesing, 2011) ซึ่งเป็นส่วนสำคัญในการช่วยให้ผู้พัฒนาโปรแกรมเข้าถึงเพชบุรีได้ง่าย สำหรับผู้ใช้งานเพชบุรีการประกาศ (post) หรือการคอมเมนต์ (comment) เป็นตัวอย่างส่วนหนึ่งของอ็อบเจกต์ (object) บนเพชบุรี โดยในแต่ละอ็อบเจกต์นั้นจะมีเลขไอดีเป็นของตัวเอง การเข้าถึงอ็อบเจกต์นั้นต้องใช้เอชทีทีพี (HTTP) ร้องขอไปยังเอนด์พอยท์ (endpoint) โดยที่เอนด์พอยท์ทำหน้าที่รับและส่งการร้องขอจากเอชทีทีพี ในการเข้าถึงข้อมูลนั้นจะถูกแสดงด้วยเลขไอดีและทาง เช็พเวอร์ของเพชบุรีจะส่งข้อมูลประเภทเจสัน (Java Script Object Notation: JSON) กลับมา

## 3 การประเมินผลการจำแนก

การหาค่าความถูกต้องแม่นยำเป็นการ ประเมินหาประสิทธิภาพของการจำแนกข้อมูล ตาม

หลักการค้นคืนสารสนเทศ สมการที่ (8) แสดงการหาค่าความถูกต้องแม่นยำ

$$\text{Accuracy} = \left( \frac{TP+TN}{TP+FN+TN+FP} \right) 100 \quad (8)$$

โดยที่ TP คือตัวอย่างข้อมูลที่อยู่ในกลุ่ม vj และตัวจำแนกทำนายว่าอยู่ในกลุ่ม vj TN คือกลุ่มตัวอย่างที่ไม่อยู่ในกลุ่มของ vj และตัวจำแนกทำนายว่าไม่อยู่ในกลุ่มของ vj FP คือตัวอย่างข้อมูลที่ไม่อยู่ในกลุ่ม vj แต่ตัวจำแนกทำนายว่าอยู่ในกลุ่ม vj และ FN คือตัวอย่างกลุ่มข้อมูลที่อยู่ในกลุ่มของ vj แต่ตัวจำแนกทำนายว่าไม่อยู่ในกลุ่มของ vj

### งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องกับการแนะนำสถานที่ การใช้หลักวิธีการจำแนกประเภทเบสส์อย่างง่าย การเชื่อมต่อกับเฟซบุ๊กด้วยการใช้งานส่วนต่อประสานโปรแกรมประยุกต์ และการประเมินประสิทธิผลของการจำแนก เพื่อนำมาใช้เป็นแนวทางในการทำงานวิจัยและช่วยเพิ่มประสิทธิภาพให้กับงานวิจัย

Wang et al. (2012) นำเสนอระบบแนะนำด้วยการใช้ข้อมูลทางประชากรศาสตร์จากเว็บไซต์ทริปแอดไวเซอร์การประยุกต์ใช้ขั้นตอนวิธีการแนะนำข้อมูลจากการวิเคราะห์ข้อมูลประชากร เพื่อทำนายการประเมินค่าความนิยมสถานที่ท่องเที่ยวของนักท่องเที่ยว โดยให้ความสนใจที่นิวยอร์ก ปารีส ลอนดอน โรม ซิดนีย์ และเบอร์ลิน โดยเลือกใช้กลุ่มข้อมูลนักท่องเที่ยวจากเว็บไซต์ทริปแอดไวเซอร์ ที่มีการประเมินค่าความนิยมสถานที่ท่องเที่ยวประเภทพิพิธภัณฑ์ สวนสัตว์ และพิพิธภัณฑ์สัตว์น้ำ งานวิจัยนี้เตรียมข้อมูลสำหรับการทดลองจากกลุ่มข้อมูลนักท่องเที่ยวจากเว็บไซต์ทริปแอดไวเซอร์ที่มีการระบุข้อมูลด้านประชากรไว้ ข้อมูลประชากรประกอบด้วยอายุ เพศ รูปแบบการท่องเที่ยว จุดประสงค์ของการท่องเที่ยว วันหยุด และผู้ร่วม

เดินทาง โดยใช้หลักการจำแนกประเภทของสามวิธีคือ หลักการจำแนกประเภทเบสส์อย่างง่าย หลักการจำแนกข่ายงานแบบเบสส์ และซัพพอร์ตเวกเตอร์แมชชีนมาเทียบกัน ผลลัพธ์ที่ได้โดยการประเมินจากค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) ผลลัพธ์ค่าเฉลี่ยความคลาดเคลื่อนแสดงให้เห็นว่าหลักวิธีการของซัพพอร์ตเวกเตอร์แมชชีนและการจำแนกประเภทเบสส์อย่างง่ายมีค่าที่ดีกว่าหลักการจำแนกข่ายงานแบบเบสส์

F. Müller and F. Thiesing, (2011) นำเสนอการนำเสนอต่อประสานโปรแกรมประยุกต์หรือเอพีไอ (application program interface: API) ที่เชื่อมต่อกับเฟซบุ๊ก (Facebook) มาใช้งานกับองค์กร มีการอธิบายส่วนประกอบต่าง ๆ ของเฟซบุ๊ก เช่น หน้าโปรไฟล์เฟซบุ๊ก กลุ่มเฟซบุ๊ก เฟซบุ๊กแฟนเพจ และเฟซบุ๊กแอป เป็นต้น อีกทั้งได้นำเสนอการใช้งานชุดเครื่องมือสำหรับพัฒนาซอฟต์แวร์บนเฟซบุ๊กซึ่งสามารถใช้งานหลากหลายภาษาในการพัฒนา

นฤพนธ์และจักรกฤษณ์ (2013) ได้นำเสนอระบบจัดการหมวดหมู่ให้เว็บไซต์การท่องเที่ยวของประเทศไทยโดยใช้หลักวิธีการจำแนกประเภทเบสส์อย่างง่ายที่มีการจำแนกตามออนโทโลยีที่ได้ออกแบบไว้ ทั้งนี้เพื่อช่วยในการเข้าถึงข้อมูลของผู้ใช้งานได้มีความสะดวกมากขึ้นด้วยการเข้าใช้งานเพียงเว็บเดียว ประสิทธิภาพของการจำแนกด้วยหลักวิธีการจำแนกประเภทเบสส์อย่างง่ายจากระบบที่พัฒนาขึ้นพบว่ามีค่าความแม่นยำ ค่าระลอก และค่า F-Measure เท่ากับ 72.60 70.99 และ 71.61% ตามลำดับ

เดชและพยุง (2554) นำเสนอวิธีการจำแนกข้อมูลด้วยวิธีร่วมกันตัดสินใจโดยใช้เทคนิคโครงข่ายประสาทเทียมร่วมกับการรวมกลุ่มตัดสินใจและมีการนำเทคนิคเอดาบูท (Adaboost) มาใช้เพื่อให้ได้ความถูกต้องในการจำแนกข้อมูลที่มีค่าสูงขึ้น โดยขึ้น

แรกนำข้อมูลตัวอย่างมาปรับให้อยู่ในรูปแบบเดียวกัน และนำเสนอวิธีการสร้างตัวจำแนกโดยใช้วิธีการสุ่มกลุ่มตัวอย่างเพื่อสร้างความแตกต่างให้กับกลุ่มข้อมูลสอน หลังจากนั้นนำกลุ่มข้อมูลสอนไปสอนให้กับตัวจำแนก ข้อมูลที่ใช้เทคนิคโครงข่ายประสาทเทียม ในการประเมินผลงานวิจัยนี้ใช้ตัววัดความถูกต้องแม่นยำ (accuracy)

จากงานวิจัยที่เกี่ยวข้องตามที่กล่าวมา ได้นำมาเป็นแนวทางในการทำงานวิจัยคือ งานวิจัยที่ (Y. Wang et al., 2012) การนำข้อมูลทางประชากรศาสตร์มาใช้ในการจำแนกประเภทประชากร โดยงานวิจัยที่นำเสนอได้นำข้อมูลโพรไฟล์จากเฟซบุ๊ก มาใช้สำหรับการจำแนกประเภท โดยนำวิธีการการเชื่อมต่อเฟซบุ๊กเพื่อเรียกใช้งานข้อมูลจากเฟซบุ๊กที่นำเสนอไว้ใน F. Müller and F. Thiesing, (2011) จากนั้นได้นำแนวทางในการใช้หลักวิธีการจำแนกประเภทเบสอย่างง่ายมาใช้กับงานวิจัยที่นำเสนอจากงานวิจัยของนฤพนธ์และจักรกฤษณ์ (2013) และได้นำแนวทางการประเมินผลด้วยตัววัดค่าความถูกต้องแม่นยำจากงานวิจัยของเดชและพุง (2554) มาใช้กับงานวิจัยที่นำเสนอ

## การสร้างกฎสำหรับการจำแนกด้วยหลักการจำแนกประเภทเบสอย่างง่าย

การสร้างกฎจากหลักการจำแนกประเภทเบสอย่างง่ายแบ่งออกเป็น 5 ส่วนคือ การศึกษาเกี่ยวกับงานวิจัย การเก็บรวบรวมข้อมูล การเรียนรู้เบสอย่างง่าย การจำแนกประเภทประชากร และการประเมินผลการจำแนก โดยภาพรวมของวิธีดำเนินการวิจัยแสดงตามรูปที่ 1

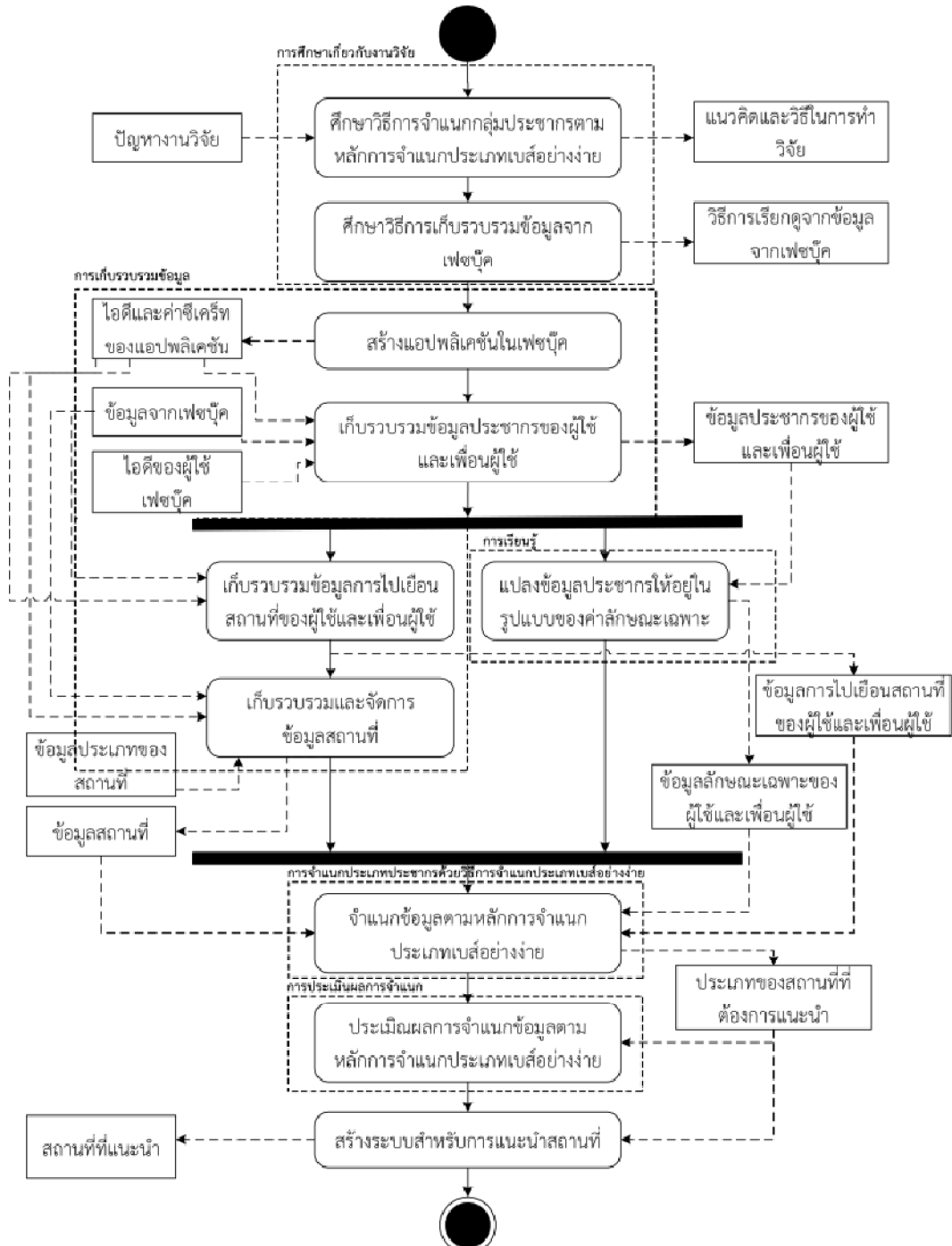
## 1 การศึกษาเกี่ยวกับงานวิจัย

การศึกษาเกี่ยวกับงานวิจัยนี้แบ่งออกเป็น 2 ส่วน คือการศึกษาวิธีการจำแนกกลุ่มประชากรตามหลักการจำแนกประเภทเบสอย่างง่าย ซึ่งในส่วนนี้จะทราบถึงแนวคิดในการสร้างกฎสำหรับการจำแนกประเภทประชากร และส่วนต่อมาคือการศึกษาวิธีการเก็บรวบรวมข้อมูลจากเฟซบุ๊ก เพื่อทราบถึงวิธีการได้มาของข้อมูล เพื่อนำมาใช้สำหรับการเรียนรู้ของหลักการจำแนกประเภทเบสอย่างง่าย

## 2 การเก็บรวบรวมข้อมูล

ในขั้นตอนนี้เป็นส่วนของการเก็บรวบรวมข้อมูลเพื่อนำไปใช้เป็นข้อมูลสอนให้กับการเรียนรู้ของหลักการจำแนกประเภทเบสอย่างง่าย โดยเก็บข้อมูลโพรไฟล์ของประชากรจากผู้ใช้เฟซบุ๊กจำนวน 600 ราย ข้อมูลโพรไฟล์ของผู้ใช้แต่ละรายประกอบด้วย 7 รายการ ได้แก่ เพศ อายุ สถานะความสัมพันธ์ ระดับการศึกษา สถานะการทำงาน ภูมิภาค และที่อยู่ ปัจจุบัน ผู้ใช้ที่ถูกเก็บรวบรวมจะต้องมีข้อมูลโพรไฟล์ครบทั้ง 7 รายการ

เนื่องจากข้อมูลโพรไฟล์ของประชากรทั้ง 7 รายการนี้เป็นปัจจัยที่มีผลต่อการเลือกสถานที่ ตัวอย่างเช่น ในปัจจัยเรื่องเพศ ลักษณะการเข้าศูนย์การค้าของผู้ชายและผู้หญิงต่างกัน ผู้ชายอาจจะเลือกเข้าร้านชุดแต่งรถยนต์ ส่วนผู้หญิงอาจจะเลือกเข้าร้านเครื่องสำอางค์ หรือในส่วนของปัจจัยเรื่องสถานะการทำงาน ผู้หญิงที่อยู่บ้านเป็นแม่บ้านอาจจะมึลักษณะการท่องเที่ยวที่แตกต่างจากผู้หญิงที่เป็นวัยทำงาน หรือลักษณะของการใช้ชีวิตประจำวันของผู้ใช้ที่อยู่กรุงเทพฯ และจังหวัดอื่น ๆ อาจจะต่างกัน ทั้งนี้เนื่องจากปัจจัยเหล่านี้จะส่งผลต่อความชื่นชอบ หรือลักษณะการดำเนินชีวิตในแต่ละวัน ซึ่งจะมีผลต่อการเลือกสถานที่ที่แตกต่างกัน



รูปที่ 1 ภาพรวมวิธีดำเนินงานวิจัย

### 3 การเรียนรู้เบื้องต้นอย่างง่าย

ขั้นตอนการเรียนรู้เป็นขั้นตอนที่ทำให้ได้ข้อมูลเพื่อนำไปใช้สำหรับการจำแนกประเภท ข้อมูล

โพรไฟล์ของประชากรที่นำมาใช้สำหรับการจำแนกประเภทจะถูกแปลงค่าเป็นมาตรวัดนามบัญญัติ (nominal scale) โดยการใช้ตัวเลขหรือตัวอักษรแทน

ค่าข้อมูลโพรไฟล์ประชากรแต่ละประเภท แสดงตาม ตารางที่ 1 และในการระบุกลุ่มให้กับกลุ่มประชากร ตัวอย่างนั้นพิจารณาจากความถี่ของการไปเยือน สถานที่จำนวน 35,084 รายการ

#### 4 การจำแนกประเภทประชากรด้วยวิธีการจำแนกประเภทเบสอย่างง่าย

ในงานวิจัยนี้จะจำแนกกลุ่มของประชากร ตามหมวดหมู่ของสถานที่ได้ระบุไว้ในเพชบู้ค 4 หมวดหมู่ คือ ร้านอาหาร โรงแรม สถานที่ท่องเที่ยว กลางแจ้ง และศูนย์การค้าและค้าปลีก พร้อมทั้งได้ กำหนดรายการหมวดหมู่ย่อยของแต่ละหมวดหมู่ ดังต่อไปนี้

หมวดหมู่ร้านอาหารประกอบด้วย อาหารไทย อาหาร สากล/นานาชาติ ของหวาน คาเฟ่ และอื่น ๆ

หมวดหมู่โรงแรมประกอบด้วย มากกว่าหรือเท่ากับ 3 ดาว 4 ดาว มากกว่าหรือเท่ากับ 5 ดาว

หมวดหมู่สถานที่ท่องเที่ยวกลางแจ้ง ประกอบด้วย ทะเล แม่น้ำ ภูเขา อุทยานแห่งชาติ และอื่น ๆ

ศูนย์การค้าและค้าปลีกประกอบด้วย การปรับปรุงบ้าน ของโบราณ คอมพิวเตอร์และอุปกรณ์ ไฟฟ้า เครื่องสำอางค์ ธนาคาร บริการอุปกรณ์การเดินทาง ร้านกล้อง ร้านเกี่ยวกับดนตรี ร้านของขวัญ ร้านค้าที่ลดราคา ร้านค้าในเขต ร้านค้าปลีก ขนาดใหญ่ ร้านค้าส่ง ร้านค้าสัตว์เลี้ยง ร้านจักรยาน ร้านซักผ้า ที่ซักด้วยตนเอง ร้านซักรีด ร้านดอกไม้ ร้านดีวีดีและวีดีโอ ร้านโทรศัพท์มือถือ ร้านป้ายสัญญาณและธง ร้านปิ่นร้านพลู ร้านเฟอร์นิเจอร์ ร้านยาร้านแว่น ร้านไวน์ ร้านสะดวกซื้อ ร้านสินค้าอุปกรณ์กีฬา ร้านเสื้อผ้า ร้าน

หนังสือ ร้านหนังสือ การ์ตูน ร้านอาหารและขายของชำ ร้านอุปกรณ์กลางแจ้ง ร้านเครื่องประดับ วัสดุงานศิลปะและงานฝีมือ ศูนย์การค้า ศูนย์การค้าขนาดใหญ่ ศูนย์รวมงานสร้างสรรค์ อุปกรณ์สำนักงาน และอื่น ๆ

ในการแบ่งแยกหมวดหมู่ย่อยให้กับสถานที่นั้นจะมีการพิจารณาจากหมวดหมู่ย่อยของสถานที่ที่ถูกระบุมาจากเพชบู้ค กฎที่สร้างขึ้นสำหรับการจำแนกโดยใช้วิธีการจำแนกประเภทเบสอย่างง่ายนั้นจะใช้สมการที่ (6) ด้วยการคำนวณจากค่าโพรไฟล์ทั้งหมด 7 รายการ และจะถูกจำแนกตามหมวดหมู่ออกเป็น 4 หมวดหมู่

#### 5 การประเมินผลการจำแนก

ในการประเมินผลการจำแนกได้ทำการทดลองกฎที่สร้างขึ้น โดยแบ่งชุดข้อมูลประชากรออกเป็น 2 ส่วน คือส่วนที่เป็นชุดข้อมูลสอน และส่วนของชุดข้อมูลทดสอบ จำนวน 480 และ 120 ตามลำดับ จากนั้นทำการทดสอบด้วยกฎที่สร้างขึ้น ค่าความแม่นยำที่ได้โดยแบ่งตามหมวดหมู่ของสถานที่ 4 หมวดหมู่ ได้แก่ร้านอาหาร โรงแรม สถานที่ท่องเที่ยว กลางแจ้งและศูนย์การค้าและค้าปลีก เป็น 78.7 72.5 67.4 และ 76.7% ตามลำดับ ข้อมูลโพรไฟล์ของชุดข้อมูลสอนและชุดข้อมูลทดสอบแสดงตามตารางที่ 1 จากการสรุปผลค่าความแม่นยำในการจำแนกประเภทแต่ละหมวดหมู่ ค่าความแม่นยำในหมวดหมู่ของร้านอาหารให้ค่าความแม่นยำที่สูงที่สุด จากการวิเคราะห์ผลข้อมูลจะเห็นได้ว่าข้อมูลสอนที่อยู่ในกลุ่มของหมวดหมู่ร้านอาหาร และสถานที่ที่อยู่ในหมวดหมู่ร้านอาหารนั้นมีจำนวนสูงสุด ดังนั้นทำให้ทราบได้ว่าปริมาณของชุดข้อมูลสอนนั้นเป็นปัจจัยที่มีผลต่อการเรียนรู้ของระบบด้วยเช่นกัน



ตารางที่ 1 ข้อมูลโพรไฟล์ นามบัญญัติ และจำนวนข้อมูลสอนและข้อมูลทดสอบ

ข้อมูลโพรไฟล์	การแปลงเป็นนามบัญญัติ	ข้อมูลสอน	ข้อมูลทดสอบ
วันเกิด	1: ช่วงอายุ 0 – 22 ปี	23	6
	2: ช่วงอายุ 23 – 60 ปี	457	114
	3: อายุมากกว่า 60 ปี	0	0
เพศ	1: ชาย	124	30
	2: หญิง	356	90
สถานะความสัมพันธ์	1: โสด	110	24
	2: ความสัมพันธ์อื่นๆ	232	73
	3: ไม่ระบุ	138	23
ระดับการศึกษา	1: ต่ำกว่าปริญญาตรี	24	2
	2: ปริญญาตรี	348	87
	3: สูงกว่าปริญญาตรี	108	31
สถานะการทำงาน	0: ไม่ทำงาน	147	20
	1: ทำงาน	333	100
ภูมิภาค	C: ภาคกลาง	285	90
	N: ภาคเหนือ	3	0
	EN: ภาคตะวันออกเฉียงเหนือ	25	3
	E: ภาคตะวันออก	2	0
	W: ภาคตะวันตก	0	0
	S: ภาคใต้	20	3
	O: ต่างประเทศ	145	24
ที่อยู่ปัจจุบัน	1: กรุงเทพฯ	362	104
	2: อื่นๆ	108	11
	3: ไม่ระบุ	10	5

### การสร้างระบบแนะนำสถานที่

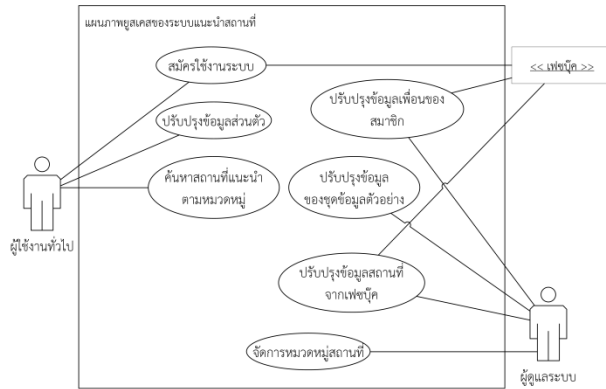
ระบบที่พัฒนาขึ้นเป็นระบบที่นำกฎที่สร้างขึ้นจากหัวข้อที่ 4 มาใช้ในการแนะนำสถานที่ให้กับผู้ใช้ แผนภาพยูสเคสของระบบแสดงตามรูปที่ 2 โดยมีการแบ่งกลุ่มผู้ใช้งานออกเป็น 2 กลุ่ม ได้แก่ กลุ่มผู้ใช้งานทั่วไปและกลุ่มผู้ดูแลระบบ

หลักการการทำงานของระบบแนะนำเพื่อแสดงผลการค้นหาสถานที่ ในส่วนของกลุ่มผู้ใช้งานทั่วไป คือ ผู้ใช้ต้องทำการสมัครใช้งานระบบ เพื่อให้ระบบดึงข้อมูลโพรไฟล์ของผู้ใช้จากเฟซบุ๊ก จากนั้นเลือกหมวดหมู่ของสถานที่ที่ต้องการ เพื่อให้ระบบ

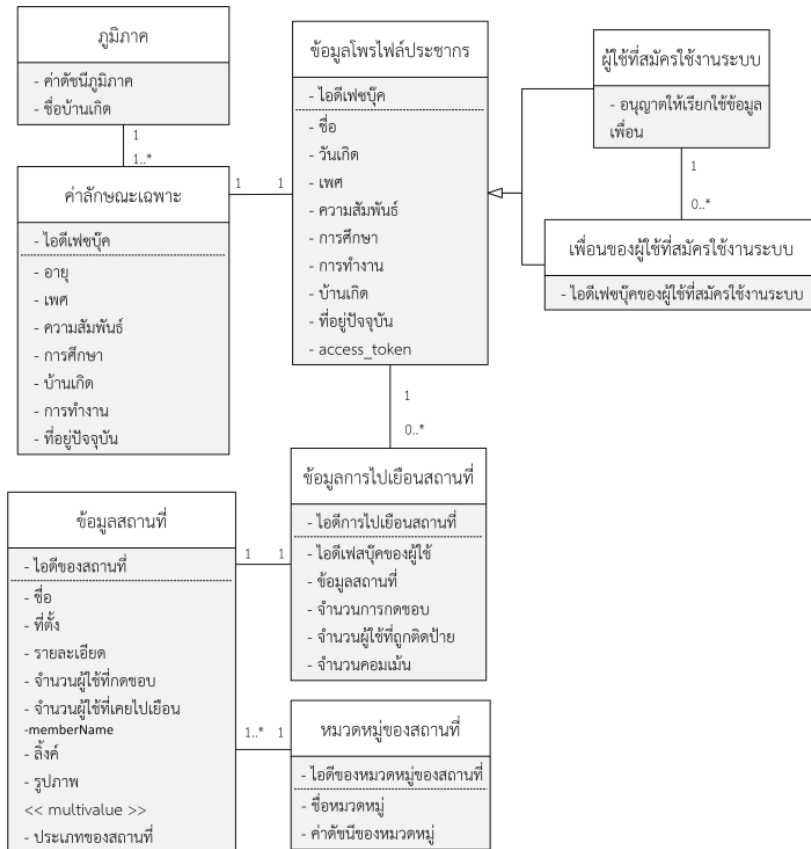
แนะนำสถานที่ในหมวดหมู่ย่อยที่เหมาะสมกับผู้ใช้ในหมวดหมู่สถานที่นั้น และระบบนำข้อมูลโพรไฟล์ของผู้ใช้ และหมวดหมู่ของสถานที่ที่ผู้ใช้เลือกมาใช้กับกฎการจำแนกที่สร้างขึ้น จากนั้นจะแสดงสถานที่ในหมวดหมู่ย่อยที่เหมาะสมกับผู้ใช้บนหน้าจอ ในส่วนของกลุ่มผู้ใช้งานทั่วไปสามารถปรับปรุงข้อมูลของผู้ใช้ให้ข้อมูลมีความทันสมัยอยู่เสมอ เพื่อใช้ในการสนับสนุนการแนะนำสถานที่ของระบบให้ได้ผลลัพธ์ที่เหมาะสมกับผู้ใช้มากที่สุด และเพื่อเพิ่มประสิทธิภาพให้กับการแนะนำสถานที่ ระบบยังมีส่วนในการขออนุญาตผู้ใช้เพื่อเรียกใช้ข้อมูลเพื่อนของผู้ใช้ เพื่อนำมาเป็นชุดข้อมูล

สอนระบบเพิ่มเติมจากข้อมูลสอนเดิมที่มีอยู่ในระบบ หลักการทำงานต่อมาคือส่วนของกลุ่มผู้ดูแลระบบ นั้น จะเน้นไปที่การปรับปรุงข้อมูลที่ใช้สำหรับการจำแนกให้ มีความทันสมัยอยู่เสมอ เพื่อให้ได้ระบบการแนะนำที่มีความทันสมัยของข้อมูลและสอดคล้องกับความสนใจใน

ปัจจุบันของผู้ใช้ จากแผนภาพยูสเคสสามารถนำมา วิเคราะห์แผนภาพคลาสได้ตามรูปที่ 3 และรูปที่ 4 แสดงตัวอย่างหน้าจอสำหรับผู้ใช้สำหรับเลือกหมวดหมู่ ของสถานที่



รูปที่ 2 แผนภาพยูสเคสของระบบแนะนำสถานที่



รูปที่ 3 แผนภาพคลาสของระบบแนะนำสถานที่



รูปที่ 4 หน้าจอสำหรับเลือกประเภทของสถานที่

จากแผนภาพคลาสในรูปแบบที่ 3 มีคลาสทั้งหมด 8 คลาส ได้แก่ คลาสข้อมูลโพรไฟล์ประชากร เก็บข้อมูลของผู้ใช้ที่นำมาเป็นข้อมูลสอนและผู้ใช้ที่สมัครใช้งานระบบ คลาสผู้ใช้ที่สมัครใช้งานระบบ เก็บรหัสอนุญาตในการเรียกใช้งานข้อมูลจากเฟซบุ๊ก เพื่อใช้ในการเชื่อมต่อกับเฟซบุ๊กคลาสเพื่อนของผู้ใช้ที่สมัครใช้งานระบบ เก็บไอดีเฟซบุ๊กของผู้ใช้งานที่อนุญาตให้นำข้อมูลเพื่อนมาใช้งาน คลาสภูมิภาค เก็บรายชื่อจังหวัดที่อยู่ในแต่ละภูมิภาค เพื่อใช้ในการแปลงค่าภูมิภาคของ用戶เป็นนามบัญญัติ คลาสคำลักษณะเฉพาะ เก็บค่านามบัญญัติที่ถูกแปลงมาจากข้อมูลโพรไฟล์ คลาสข้อมูลการไปเยือนสถานที่ เก็บรายละเอียดของการไปเยือนสถานที่ของผู้ใช้ คลาสข้อมูลสถานที่ เก็บรายละเอียดของสถานที่ และคลาสหมวดหมู่ของสถานที่ เก็บข้อมูลหมวดหมู่ของสถานที่

### สรุปผลงานวิจัยและแนวทางการดำเนินการในอนาคต

งานวิจัยนี้ได้นำเสนอวิธีการจำแนกประเภทประชากรด้วยวิธีการจำแนกประเภทเบสอย่างง่าย ด้วยการใช้ค่าค่าโพรไฟล์ของประชากรทั้งหมด 7 รายการ ได้แก่ เพศ อายุ ความสัมพันธ์ การศึกษา การทำงาน ภูมิภาคของ และที่อยู่ปัจจุบัน เป็นปัจจัยในการจำแนก โดยการเก็บรวบรวมข้อมูลประชากรจากข้อมูล

โพรไฟล์ในเครือข่ายสังคมเฟซบุ๊ก จำนวน 600 ราย โดยจำแนกตามหมวดหมู่ของสถานที่ 4 หมวดหมู่ ได้แก่ ร้านอาหาร โรงแรม สถานที่ท่องเที่ยวกลางแจ้ง และศูนย์การค้าและค้าปลีก จากความถี่ของการไปเยือนสถานที่จำนวน 35,084 รายการ และจากการทดสอบการจำแนกเพื่อวัดประสิทธิภาพจากค่าความแม่นยำของการจำแนกในแต่ละหมวดหมู่ของสถานที่ทั้ง 4 รายการ ได้แก่ ร้านอาหาร โรงแรม สถานที่ท่องเที่ยวกลางแจ้ง และศูนย์การค้าและค้าปลีก พบว่ามีค่าความแม่นยำเป็น 78.7 72.5 67.4 และ 76.7% ตามลำดับ

แนวทางการดำเนินงานในอนาคตคือ การหาปัจจัยอื่นเพื่อเพิ่มประสิทธิภาพของการจำแนกประเภทด้วยวิธีการจำแนกประเภทเบสอย่างง่าย ตัวอย่างเช่น ข้อมูลจำนวนการไปเยือนในแต่ละสถานที่ และจำนวนการให้ความเห็นโดยการกดชอบ (like) ของแต่ละสถานที่ เป็นต้น หรืออาจจะใช้ข้อมูลจากการรวบรวมค่าความพึงพอใจของผู้ใช้หลังจากการแนะนำสถานที่ เป็นอีกหนึ่งข้อมูล ในการพิจารณา ร่วมกับการจำแนกประเภทแบบเบสอย่างง่าย

### เอกสารอ้างอิง

เดช ธรรมศิริ และ พยุง มีสีจ. (2554). Adaboost Artificial Neural Network Ensemble Learning for Classification. วารสารเทคโนโลยีสารสนเทศ 7, 14(2554): 7-12.

- นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์ นมะหุต. (2013). Thailand Tourism Web Clustering System using Naïve Bayes Algorithm. The 9<sup>th</sup> National Conference on Computing and Information Technology.
- บุญเสริม กิจสิริกุล. (2548). ปัญญาประดิษฐ์ Artificial Intelligence เอกสารคำสอนวิชา 2110654. เวอร์ชัน 1.0.2, 2548.
- “Key Facts: Statistics”. [Online]. Available: <https://newsroom.fb.com/Key-Facts>, June 2013 [September 4, 2013].
- Müller, F. and Thiesing, F. (2011). Social Networking APIs for Companies An Example of using the Facebook API for Companies. International Conference on Computational Aspects of Social Networks (CASoN).
- Srivastava, J. (2008). Data Mining for Social Network Analysis. IEEE ISI 2008 Invited Talk (III).
- “Top 10 Social Media Websites”. [Online]. Available: <http://www.experian.com/hitwise/online-trends-social-media.html>. August 31, 2013 [September 6, 2013].
- Wang, Y. Chan, S. C. and Ngai, G. (2012). Applicability of Demographic Recommender System to Tourist Attractions: A Case Study on TripAdvisor. IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

