

การศึกษาเบื้องหลังผู้ใช้บริการเว็บไซต์ กรณีศึกษา: HomeDD.com
Web Usage Mining Case Study of HomeDD.com

ดร.จันทรเจ้า มงคลนาวิน *
บัณฑิต เชี่ยวเจริญ **

บทคัดย่อ

พาณิชย์อิเล็กทรอนิกส์ได้เข้ามามีบทบาทที่สำคัญต่อการทำธุรกิจในยุคปัจจุบัน ทั้งในด้านการส่งเสริมช่องทางการจำหน่ายและการสื่อสารกับลูกค้า การพัฒนารูปแบบของสินค้าและบริการใหม่ ๆ รวมทั้งการส่งเสริมภาพลักษณ์ขององค์กร นอกจากนี้ประโยชน์โดยตรงที่ได้รับ องค์กรสามารถสร้างประโยชน์ทางอ้อมจากพาณิชย์อิเล็กทรอนิกส์ด้วยการวิเคราะห์ Web Access Log หรือไฟล์บันทึกรายละเอียดการร้องขอข้อมูลจากผู้ใช้บริการเว็บไซต์ การวิเคราะห์ Web Access Log โดยเฉพาะอย่างยิ่ง การทำเหมืองข้อมูลเว็บไซต์ (Web Mining) หรือ การประยุกต์ใช้เทคนิคเหมืองข้อมูลกับ Web Access Log ส่งผลให้องค์กรสามารถทราบถึงพฤติกรรมของผู้เข้าเยี่ยมชมเว็บไซต์ และสามารถนำผลที่ได้จากการศึกษามาประยุกต์ใช้ในการบริหารจัดการเว็บไซต์ให้เหมาะสมกับความต้องการของผู้ใช้บริการ รวมทั้งอาจนำผลการศึกษาที่ได้ไปประกอบการบริหารจัดการองค์กรในด้านอื่น ๆ

ในการศึกษาครั้งนี้ ข้อมูล Web Access Log ของเว็บไซต์ HomeDD.com ระหว่างวันที่ 3 สิงหาคม 2546 ถึง 30 กันยายน 2546 ได้ถูกนำมาใช้เป็นกรณีศึกษา โดยในบทความนี้ ผู้ศึกษาได้อธิบายขั้นตอนในการทำ Web Mining กับข้อมูลที่ใช้เป็นกรณีศึกษา รายงานผลการศึกษารวมทั้งแนวทางการประยุกต์ใช้ผลที่ได้ในการบริการจัดการเว็บไซต์พาณิชย์อิเล็กทรอนิกส์

คำสำคัญ

การทำเหมืองข้อมูลเว็บไซต์ การวิเคราะห์เว็บลอค พาณิชย์อิเล็กทรอนิกส์ พฤติกรรมผู้ใช้บริการเว็บไซต์ การทำเหมืองข้อมูล

1. ความเป็นมาของการศึกษา

การศึกษานี้เกิดขึ้นเนื่องจากการที่ Web Server สามารถถูกกำหนดให้มีการจัดเก็บข้อมูลที่เรียกว่า Web Access Log โดยข้อมูลดังกล่าว โดยปกติแล้วจะมีขนาดใหญ่มากพอที่จะนำมาประยุกต์ใช้กับเทคนิคเหมืองข้อมูล ลักษณะการประยุกต์ดังกล่าวสามารถกล่าวโดยสังเขปได้ดังนี้

Web Access Log

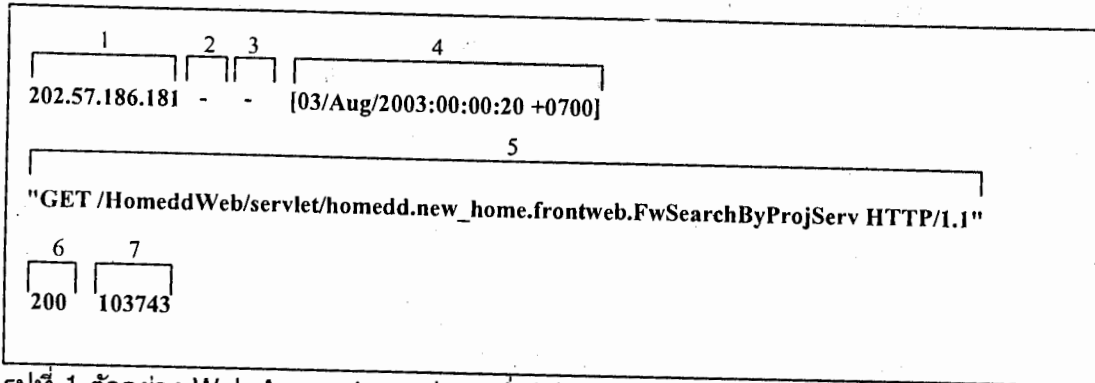
Web Access Log หรือ Web Server Log เป็นไฟล์ในเครื่องเซิร์ฟเวอร์ของเว็บไซต์ (Web Server) ที่ใช้เก็บรายละเอียดการร้องขอข้อมูลจากเครื่องของผู้ใช้บริการ (Web Client) โดยผู้ดูแล Web Server สามารถกำหนดให้มีการบันทึกข้อมูลใน Web Access Log เมื่อมีการติดตั้ง Web Server

Web Server จะทำการบันทึกข้อมูลทุกครั้งเมื่อได้รับการร้องขอข้อมูล (Access Request) ไม่ว่า Access Request นั้นจะได้รับการตอบสนอง (Response) หรือไม่ Web Access log มี 2 รูปแบบ คือ รูปแบบทั่วไป (Common Log Format) และรูปแบบผสม (Combine Log Format) [2]

1. รูปแบบทั่วไป (Common Log Format)

เป็นรูปแบบซึ่งกำหนดเป็นมาตรฐานใน Web Server มีลักษณะดังตัวอย่างในรูปที่ 1 โดยแต่ละกลุ่มข้อมูลในรูปที่ 1 มีความหมายดังแสดงในตารางที่ 1

* อาจารย์ประจำสาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
** ตำแหน่ง Assistant Manager ฝ่าย System Integration and Operation แผนก MIS and Internal System Development บริษัทพันทวิช จำกัด



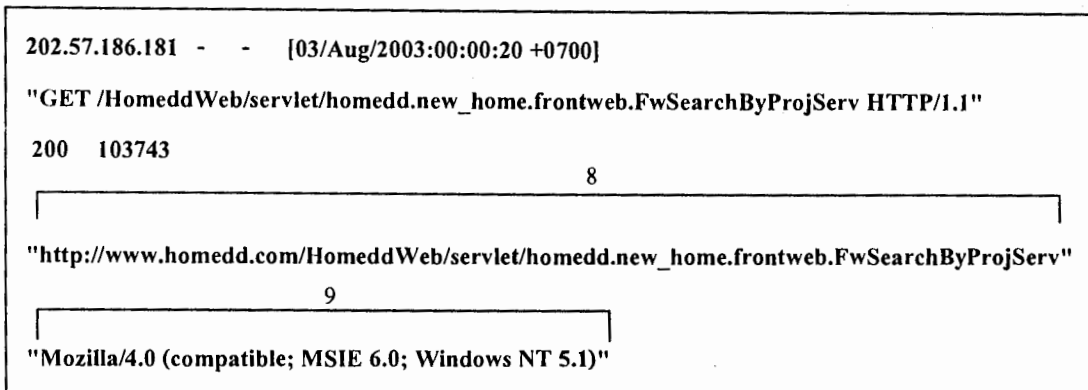
รูปที่ 1 ตัวอย่าง Web Access Log รูปแบบทั่วไป

กลุ่ม	ความหมาย
1	แสดงข้อมูลหมายเลข IP Address ของ Web Client
2	แสดงชื่อเครื่องที่ใช้ในการส่ง Access Request
3	แสดงรหัสของผู้ใช้ (User ID) ที่ส่ง Access Request
4	แสดงข้อมูลวันและเวลาที่มีการส่ง Access Request
5	แสดงข้อมูลเกี่ยวกับไฟล์ที่ได้รับการร้องขอใน Access Request
6	แสดงข้อมูลการตอบสนอง (Response) ของ Web Server ต่อ Access Request ในกรณีที่ตำแหน่งที่หนึ่งเป็นเลข 1, 2 และ 3 หมายความว่าสถานะที่ Access Request ได้รับการตอบสนองอย่างสมบูรณ์ ในกรณีที่ตำแหน่งที่หนึ่งเป็นเลข 4 และ 5 แสดงถึงการที่ Web Server ไม่สามารถตอบสนอง Access Request ได้
7	แสดงขนาดเป็นจำนวนไบต์ของไฟล์ที่ Web Server ตอบสนองต่อ Access Request

ตารางที่ 1 ความหมายของข้อมูลใน Web Access Log รูปแบบทั่วไป

2. รูปแบบผสม (Combine Log Format)

Web Access Log รูปแบบผสมจะจัดเก็บข้อมูลในส่วนต้น (กลุ่มที่ 1 - 7) เหมือนกับ Web Access Log ในรูปแบบทั่วไป แต่จะมีการจัดเก็บข้อมูลในส่วนท้ายเพิ่มขึ้น 2 กลุ่มคือ กลุ่มที่ 8 และ 9 โดยมีลักษณะดังตัวอย่างในรูปที่ 2



รูปที่ 2 ตัวอย่าง Web Access Log รูปแบบผสม

ข้อมูลกลุ่มที่ 8 เรียกว่า "Referrer" แสดงข้อมูลเกี่ยวกับหน้าเว็บไซต์ที่ผู้ใช้ใช้ส่ง Access Request มายัง Web Server ในลักษณะของการใช้ link เชื่อมโยงระหว่างเว็บไซต์

ข้อมูลกลุ่มที่ 9 เรียกว่า "User-Agent" แสดงรายละเอียดของบราวเซอร์¹ ของเครื่องคอมพิวเตอร์ที่ผู้ใช้ส่ง Access Request มายัง Web Server ซึ่งเป็นข้อมูลชนิด เวอร์ชันของบราวเซอร์ และระบบปฏิบัติการของเครื่องคอมพิวเตอร์ที่เป็น Web Client

Web Mining

Web Mining หรือ การทำเหมืองข้อมูลของเว็บไซต์ หมายถึง การใช้เทคนิคการทำเหมืองข้อมูลเพื่อค้นหาตัวแบบจากเอกสารที่เป็นเว็บเพจและ การให้บริการของเว็บไซต์ Web Mining สามารถแบ่งออกเป็น 2 กลุ่ม คือ [6, 7]

1. การทำเหมืองข้อมูลเนื้อหาเว็บไซต์ (Web-Content Mining)
2. การทำเหมืองข้อมูลการใช้งานเว็บไซต์ (Web-Usage Mining)

1. การทำเหมืองข้อมูลเนื้อหาเว็บไซต์ (Web-Content Mining)

เป็นการทำเหมืองข้อมูลเพื่อศึกษาความสัมพันธ์ของเนื้อหาและโครงสร้างการเชื่อมโยงระหว่างหน้าเว็บ (Web Page) ในเว็บไซต์ โดยมีเป้าหมายที่จะเพิ่มความสะดวกในการค้นหาข้อมูล (Finding) และการกรองข้อมูล (Filter) ให้ตรงกับความต้องการของผู้ใช้บริการเว็บไซต์มากที่สุด โดยมีตัวอย่างการประยุกต์ใช้ผลที่ได้ดังนี้

- ผู้ให้บริการจะนำเสนอเนื้อหาในหน้าเว็บให้ตรงกับความต้องการของผู้ใช้ที่กำหนดไว้ก่อนแล้ว (User Profile) โดยผู้ให้บริการต้องสมัครเป็นสมาชิกของเว็บไซต์นั้น และทำการกำหนดสิ่งที่ผู้ให้บริการสนใจเก็บไว้เป็น User Profile เมื่อมีข้อมูลข่าวสารใหม่ ผู้ให้บริการจะทำการตรวจสอบสิ่งที่ผู้ให้บริการกำหนดไว้กับข้อมูลข่าวสารนั้น หากตรงกันจะทำการแสดงผลให้ผู้ให้บริการทราบ โดยใช้เทคนิคเหมืองข้อมูลในกระบวนการจับคู่ระหว่างเนื้อหาและความต้องการของผู้ใช้
- ผู้ให้บริการจะทำการแบ่งหมวดหมู่ของเนื้อหาในเว็บไซต์ให้เหมาะสมกับความต้องการของกลุ่มผู้ชมเว็บไซต์ ซึ่งผู้ให้บริการต้องลงทะเบียนกับเว็บไซต์ โดยทางผู้ให้บริการจะใช้เทคนิคเหมืองข้อมูลเนื้อหาเพื่อการแบ่งหมวดหมู่ของเนื้อหาตามลักษณะการ Access ของผู้ใช้

2. การทำเหมืองข้อมูลการใช้งานเว็บไซต์ (Web-Usage Mining)

เป็นการทำเหมืองข้อมูลเพื่อต้องการทราบพฤติกรรมของผู้ใช้บริการเว็บไซต์ เพื่อวัตถุประสงค์สำคัญ 2 ประการ คือ

1. เพื่อทราบถึงพฤติกรรมกรรมการเยี่ยมชมของผู้ใช้บริการ และปรับปรุงโครงสร้างเว็บไซต์ให้ตอบสนองต่อความต้องการในการเข้าชมเว็บไซต์ของผู้ใช้
2. เพื่อเพิ่มความสะดวกรวดเร็วทั้งในการค้นหาและการทำรายการของผู้ใช้บริการเว็บไซต์

โดยแหล่งข้อมูลที่สำคัญสำหรับการทำเหมืองข้อมูลประเภทที่สองนี้ คือ Web Access Log ตัวอย่างการประยุกต์ใช้ผลที่ได้จากการทำ Web-Usage Mining มีดังนี้

- Statistic Analysis การวิเคราะห์ข้อมูลทางสถิติเพื่อปรับปรุงระบบ เช่น ความถี่ของจำนวนหน้าเว็บที่มีการเรียกดู และค่าเฉลี่ยของจำนวนหน้าเว็บที่มีการเรียกดูในการเยี่ยมชมแต่ละครั้ง
- Clustering การจัดกลุ่มของผู้ใช้บริการด้วยความคล้ายคลึงของลักษณะการเรียกดูข้อมูล หรือการจัดกลุ่มหน้าเว็บตามพฤติกรรมกรรมการเรียกดูของผู้ใช้บริการ
- Classification การหารูปแบบการซื้อของผู้ใช้บริการ ซึ่งบางครั้งต้องอาศัยข้อมูลจากฐานข้อมูลที่เกี่ยวข้องของประกอบการศึกษา
- Association เป็นการหาความสัมพันธ์ของหน้าเว็บเพื่อจัดรูปแบบการนำเสนอ Web Page ให้สอดคล้องกับความสนใจของผู้ใช้เว็บไซต์

¹ โปรแกรมที่อำนวยความสะดวกในการติดต่อ Web server รับข้อมูลและแปลข้อมูลเพื่อแสดงผลบนหน้าจอของผู้ใช้ เช่น โปรแกรม Internet Explorer และโปรแกรม Netscape Communicator

2. วัตถุประสงค์ของการศึกษา

เพื่อศึกษาวิธีการวิเคราะห์พฤติกรรมผู้ใช้บริการเว็บไซต์ (เฉพาะกลุ่มผู้เยี่ยมชมเท่านั้น) ด้วยเทคนิคเหมืองข้อมูลและเสนอแนะแนวทางการประยุกต์ใช้ผลลัพธ์ตามหลักการบริหารความสัมพันธ์กับลูกค้าและการบริหารเว็บไซต์ โดยข้อมูล Web Access Log ของเว็บไซต์ HomeDD.Com เป็นกรณีศึกษา

3. วิธีการศึกษา

3.1 สมมติฐานเบื้องต้นของการศึกษา

เนื่องจากการเข้าใช้บริการ Web Site ครั้งหนึ่ง ๆ Access Request อาจไม่ถูกส่งไปยัง Web Server ทุกครั้งที่ Web Client ร้องขอข้อมูล โดยเฉพาะอย่างยิ่งในกรณีที่มีการติดตั้ง Proxy Server² ในฝั่งของผู้ใช้บริการเว็บไซต์ กรณีเช่นนี้ส่งผลให้ข้อมูลใน Web Access Log ไม่ครบถ้วน ซึ่งส่งผลให้ Web Access Log ไม่สะท้อนพฤติกรรมการใช้บริการเว็บไซต์ที่แท้จริง ดังนั้นจึงจำเป็นต้องกำหนดสมมติฐานเบื้องต้นของการศึกษาว่า ผู้ใช้เข้าใช้บริการเว็บไซต์ด้วยเครื่องคอมพิวเตอร์ที่เชื่อมโยงกับเครือข่ายอินเทอร์เน็ตโดยตรงโดยไม่ผ่าน Proxy Server

3.2 ข้อมูลที่ใช้ในการศึกษา

ข้อมูลที่ใช้ในการศึกษาเป็น Web Access Log ของเว็บไซต์ HomeDD.com [3] ซึ่งเป็น Web Access Log ในรูปแบบผสม (Combine Log Format) ที่เก็บรวบรวมในช่วงระยะเวลาตั้งแต่วันที่ 3 สิงหาคม 2546 ถึงวันที่ 30 กันยายน 2546 โดยมีข้อมูลเริ่มต้นทั้งหมด 1,927,570 รายการ

ในช่วงเวลาดังกล่าว เว็บไซต์ HomeDD.com มีการให้บริการข้อมูลหลัก 10 กลุ่ม คือ

- ข้อมูลสำหรับผู้ต้องการบ้านใหม่ (New Home)
- ข้อมูลประกาศซื้อขายบ้าน (Home Classified)
- ข้อมูลเช่าที่อยู่อาศัย (Home Rental)
- ข้อมูลคอนโดมิเนียม (Condominium)
- บริการขอสินเชื่อ (Home Finance)
- ข้อมูลเกี่ยวกับการซ่อมแซมและตกแต่ง (Home Service)
- ข้อมูลเกี่ยวกับการซ่อมแซมด้วยตนเอง (Do It Yourself)
- นิตยสารเกี่ยวกับบ้าน (Home Magazine)
- ข้อมูลเกี่ยวกับการตกแต่งสวน (Home Garden)
- ข้อมูลเกี่ยวกับการตกแต่งบ้าน (Home Décor)

² Proxy Server เป็นเครื่อง Server เพื่อเพิ่มประสิทธิภาพในการสื่อสารระหว่างองค์กรกับเครือข่ายอินเทอร์เน็ต ซึ่งมักมีการติดตั้งในองค์กรที่มีขนาดใหญ่ โดยทุก ๆ Access request ที่ส่งออกไปจะถูกส่งผ่าน Proxy server ก่อนเสมอ ในทางตรงกันข้ามข้อมูลที่ถูกส่งกลับมาจาก Web server ก็จะถูกส่งผ่าน Proxy server ก่อนเช่นเดียวกัน โดย Proxy server จะบันทึกข้อมูลที่รับจาก Web server บางส่วนไว้ที่ตัวเอง ในกรณีที่ผู้ใช้เรียกดูข้อมูลจากเว็บไซต์ที่ Proxy server มีข้อมูลจัดเก็บอยู่ Proxy server ก็จะไม่ทำการ request ข้อมูลไปยัง Web server แต่จะจัดส่งข้อมูลที่มีอยู่แล้วให้กับเครื่องคอมพิวเตอร์ของผู้ใช้โดยตรง ดังนั้นในระบบที่มีการติดตั้ง Proxy server Access request ทุกอันที่ส่งออกไปจะถูกส่งในนามของ Proxy server

3.3 เครื่องมือที่ใช้ในการศึกษา

ซอฟต์แวร์ที่ใช้ในการศึกษาคือ ชุดโปรแกรมสำเร็จรูป Enterprise Miner ของ SAS System [5] โดยเทคนิคเหมืองข้อมูล 2 เทคนิคที่ใช้ในการศึกษาคือ

1. เทคนิค Clustering

การทำเหมืองข้อมูลเว็บไซต์ด้วยเทคนิค Clustering เป็นเทคนิคที่ใช้ในการจัดกลุ่มข้อมูลที่มีแนวโน้มหรือลักษณะคล้ายคลึงกันไว้ด้วยกัน โดยไม่ได้มีการระบุกลุ่มข้อมูลไว้ล่วงหน้า ซึ่งจัดอยู่ในลักษณะ Unsupervised Learning

2. เทคนิค Decision Tree

การทำเหมืองข้อมูลเว็บไซต์ด้วยเทคนิค Decision Tree เป็นเทคนิคแบบ Classification ซึ่งใช้ในการหาความสัมพันธ์ระหว่างลักษณะข้อมูลกับกลุ่มที่มีการกำหนดไว้ล่วงหน้า โดยข้อมูลที่ใช้ในการศึกษาได้รับการจำแนกกลุ่ม (Class) ไว้แล้วอย่างชัดเจน การศึกษาความสัมพันธ์แบบนี้จัดอยู่ในลักษณะ Supervised Learning ในการศึกษาครั้งนี้ ข้อมูลที่ใช้ในขั้นตอนนี้เป็นผลที่ได้จากการจัดกลุ่มด้วยเทคนิค Clustering และผลลัพธ์ที่ได้จากขั้นตอนนี้เป็นแผนภาพต้นไม้ (Tree) ที่ช่วยให้เข้าใจลักษณะสำคัญที่จำแนกข้อมูลแต่ละกลุ่มออกจากกัน

3.4 ขั้นตอนการศึกษา

1. เก็บรวบรวมข้อมูล Web Access log จากเว็บไซต์ HomeDD.com ในช่วงเวลาที่กำหนด
2. นำข้อมูลที่ได้มาจัดรูปแบบให้เหมาะสมกับการศึกษา
3. นำข้อมูลมาวิเคราะห์หาพฤติกรรมผู้ใช้บริการเว็บไซต์ HomeDD.com ด้วยเทคนิคเหมืองข้อมูล

ขั้นตอนที่ 2 หรือขั้นตอนการประมวลข้อมูล (Data Pre-Processing) เป็นขั้นตอนที่สำคัญเป็นอย่างยิ่งในการทำเหมืองข้อมูล เนื่องจากเป็นขั้นตอนที่จะจัดการกับข้อมูลที่ไม่สมบูรณ์ (Data Cleaning) และจัดรูปแบบของข้อมูลให้มีความหมายเหมาะสมกับการศึกษา (Data Transformation) ในการศึกษาครั้งนี้ ผู้ศึกษาได้ดำเนินการประมวลข้อมูลดังนี้

การคัดเลือกข้อมูลสำหรับการวิเคราะห์

เนื่องจากข้อมูลจาก Web Access Log นั้นเป็นข้อมูลที่เกิดจาก Access Request ที่ส่งมายัง Web Server ทั้งหมด รวมถึง Access Request สำหรับไฟล์รูปภาพ หรือ Access Request ที่เกิดจากโปรแกรมคอมพิวเตอร์ เช่น Access Request ที่เกิดจาก Search Engines ต่าง ๆ ซึ่งไม่สามารถนำมาใช้ประโยชน์ในการวิเคราะห์พฤติกรรมผู้ใช้บริการได้ ดังนั้นผู้ศึกษาจึงได้ดำเนินการคัดเลือกข้อมูลในส่วนนี้ออก ในชุดข้อมูลเริ่มต้นได้มีข้อมูล Access Request บางส่วนที่เกิดจากกลุ่มผู้ใช้ที่เป็นผู้ประกอบการและผู้ดูแลเว็บไซต์ที่เข้ามาปรับปรุงข้อมูลของตนในเว็บไซต์ HomeDD.com ซึ่งไม่ใช่กลุ่มเป้าหมายในการศึกษา ดังนั้นข้อมูลดังกล่าวจึงถูกคัดเลือกรวมออกเช่นเดียวกัน นอกจากนี้ ผู้ศึกษาได้เลือกข้อมูลที่มีการ Access Request ไปยังหน้าเว็บเพจประเภท Pop-Up³ ต่าง ๆ ออกจากชุดข้อมูล เนื่องจากในเว็บไซต์ที่ใช้เป็นกรณีศึกษานั้น หน้าเว็บเพจประเภท Pop-Up ไม่ได้มีความสัมพันธ์กับหน้าเว็บอื่นแต่อย่างใด แต่จะแสดงทุกครั้งเมื่อเข้าเยี่ยมชมเว็บไซต์ที่หน้าหลัก หลังจากทำการคัดเลือกข้อมูลที่ไม่มีความสัมพันธ์กับการวิเคราะห์ออกแล้ว ชุดข้อมูลมีขนาด 1,645,745 รายการ

³ เว็บเพจประเภท Pop-Up คือหน้าเว็บที่แสดงเป็นหน้าต่างเล็ก ๆ ที่มักจะแสดงเมื่อผู้เข้าสู่เว็บไซต์หนึ่ง ๆ โดยที่ผู้ใช้ไม่ได้ร้องขอและมักเป็นหน้าเว็บที่แสดงข้อความข่าวประชาสัมพันธ์ของเว็บไซต์นั้น ๆ

การจัดเตรียมข้อมูลสำหรับการวิเคราะห์ด้วยวิธีการทำเหมืองข้อมูล

นอกจากจะคัดเลือกข้อมูลที่ไม่มีความสัมพันธ์กับการศึกษาออกแล้ว ผู้ศึกษาได้ทำการจัดเตรียมข้อมูลให้มีความเหมาะสมสำหรับการวิเคราะห์เพื่อตอบปัญหาที่ต้องการได้ โดยมีการเตรียมข้อมูลต่างๆ ดังนี้

- กำหนดช่วงเวลาการใช้บริการ
ข้อมูลที่ได้จาก Access Log จะแสดงในรูปแบบ HH:MM:SS (ชั่วโมง:นาที:วินาที) ผู้ศึกษาได้ปรับช่วงเวลาการเข้าใช้บริการให้เป็นกลุ่มที่เหมาะสมสำหรับการวิเคราะห์ โดยในการศึกษาคั้งนี้ ผู้ศึกษาได้แบ่งช่วงเวลาการใช้บริการเป็น 4 ช่วง ดังรายละเอียดในตารางที่ 2

ช่วงเวลา	เวลาเริ่มต้น	เวลาสิ้นสุด
Early Morning	00:00:01	06:00:00
Morning	06:00:01	12:00:00
Afternoon	12:00:01	18:00:00
Night	18:00:01	24:00:00

ตารางที่ 2 เกณฑ์การจัดช่วงเวลาการใช้บริการ

- กำหนดวันเข้าใช้บริการเว็บไซต์
ข้อมูลจาก Access Log จะมีรูปแบบ DD/MMM/YYYY (วันที่/ชื่อย่อเดือน/ปี) ดังนั้นผู้ศึกษาปรับให้อยู่รูปแบบวันในสัปดาห์ เพื่อให้มีความหมายเหมาะสมสำหรับการวิเคราะห์
- กำหนดแหล่งที่มาของผู้ใช้บริการ
ข้อมูลแหล่งที่มาสามารถตรวจสอบได้จากส่วน Referrer (ข้อมูลกลุ่มที่ 8 ในรูปที่ 2) โดยจะตรวจสอบจาก Access Request ครั้งแรกที่เกิดขึ้นในการเยี่ยมชมแต่ละครั้งของผู้ใช้บริการ ในการศึกษาี้ แหล่งที่มาของผู้ใช้บริการได้ถูกแบ่งออกเป็น 3 กลุ่ม ดังรายละเอียดในตารางที่ 3

แหล่งที่มา	คำอธิบาย
HomeDD	กรณีผู้ใช้บริการเยี่ยมชมเว็บไซต์โดยตรง
Portal	กรณีผู้ใช้บริการเยี่ยมชมเว็บไซต์ด้วยการ link มาจากเว็บไซต์อื่น
Search Engine	กรณีผู้ใช้บริการเยี่ยมชมเว็บไซต์ด้วยการ link มาจากผลการค้นหาของ Search Engine

ตารางที่ 3 การจัดแหล่งที่มาของผู้ใช้บริการและคำอธิบาย

- กำหนดกลุ่มหน้าเว็บที่ผู้ใช้เยี่ยมชม
เป็นการกำหนดชื่อไฟล์ต่างๆ ที่ปรากฏใน Access Request ให้อยู่ในกลุ่ม 13 กลุ่ม ซึ่งประกอบไปด้วยกลุ่มข้อมูล 10 กลุ่มที่ HomeDD.com ให้บริการดังระบุในหัวข้อ 3.2 และกลุ่มเพิ่มเติมอีก 3 กลุ่มดังต่อไปนี้ คือ

กลุ่มที่ 11 Main	กลุ่มของหน้าเว็บเพจหลักของเว็บไซต์
กลุ่มที่ 12 Member	กลุ่มของหน้าเว็บเพจสำหรับสมาชิก
กลุ่มที่ 13 Web Board	กลุ่มของหน้าเว็บเพจสำหรับ Web Board

- กำหนดช่วงการเยี่ยมชม (Session)
จากผลการศึกษาของ www.webtrends.com ผู้ใช้เว็บไซต์มักจะใช้เวลาอยู่กับหน้าเว็บหนึ่ง ๆ ไม่เกิน 30 นาที [4] ปัจจัยที่ใช้กำหนดขอบเขตในการเยี่ยมชมคือ หมายเลข IP Address และ ระยะเวลาของเวลาที่เกิด Access Request ที่มี IP Address เดียวกัน โดยมีเงื่อนไข คือ หากระยะเวลาของเวลาที่เกิด Access Request ห่างจากเวลาที่เกิด Access Request ครั้งสุดท้ายของหมายเลข IP Address เดียวกันเกิน 30 นาที จะ

พิจารณาเป็นการเยี่ยมชมครั้งใหม่ แต่หากระยะเวลาห่างน้อยกว่า 30 นาที จะพิจารณาเป็นการเยี่ยมชมต่อเนื่องภายในช่วง (session) เดียวกัน

- การจัดรูปแบบข้อมูลสำหรับทำเหมืองข้อมูล
ผู้ศึกษาได้จัดทำตารางข้อมูล ซึ่งตารางข้อมูลจะมีคอลัมน์ประกอบด้วย รหัสการเยี่ยมชม (VISITID), หมายเลข IPAddress และชื่อกลุ่มหลักของหน้าเว็บทั้ง 13 กลุ่ม โดยในแต่ละเรคคอร์ดจะมีค่าเริ่มต้นของทุกคอลัมน์เป็น 0 และในการเยี่ยมชม 1 ช่วงเวลาที่ต่อเนื่องกัน (Session เดียวกัน) ถ้าผู้ใช้บริการเว็บไซต์เรียกดูข้อมูลในกลุ่มหน้าเว็บใด คอลัมน์ที่ตรงกับกลุ่มของหน้าเว็บที่เรียกดูจะถูกเปลี่ยนค่าเป็น 1

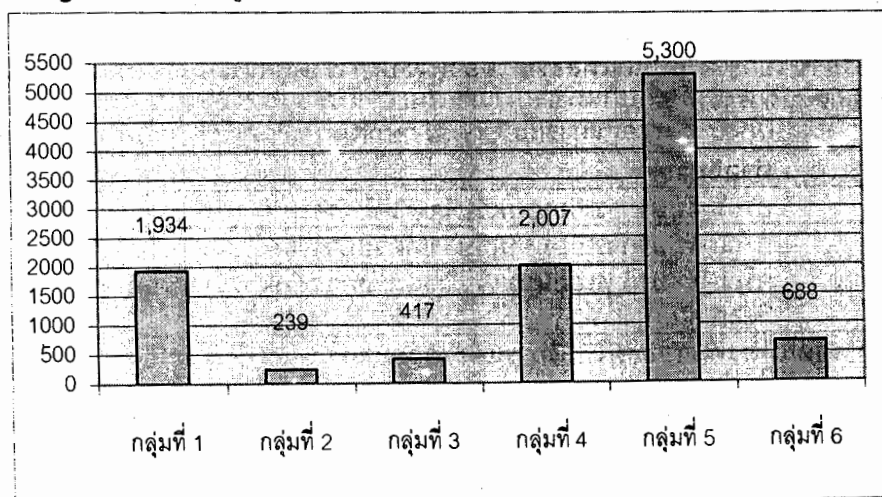
ข้อมูลที่ใช้เป็น Training set ของการทำเหมืองข้อมูลได้จากการสุ่มตัวอย่างจำนวน 500,000 รายการคิดเป็น 30.38% จากทั้งหมด 1,645,745 รายการ สรุปเป็นข้อมูลการเยี่ยมชมทั้งหมด 10,585 ครั้ง

4. ผลการศึกษาและการประยุกต์ใช้

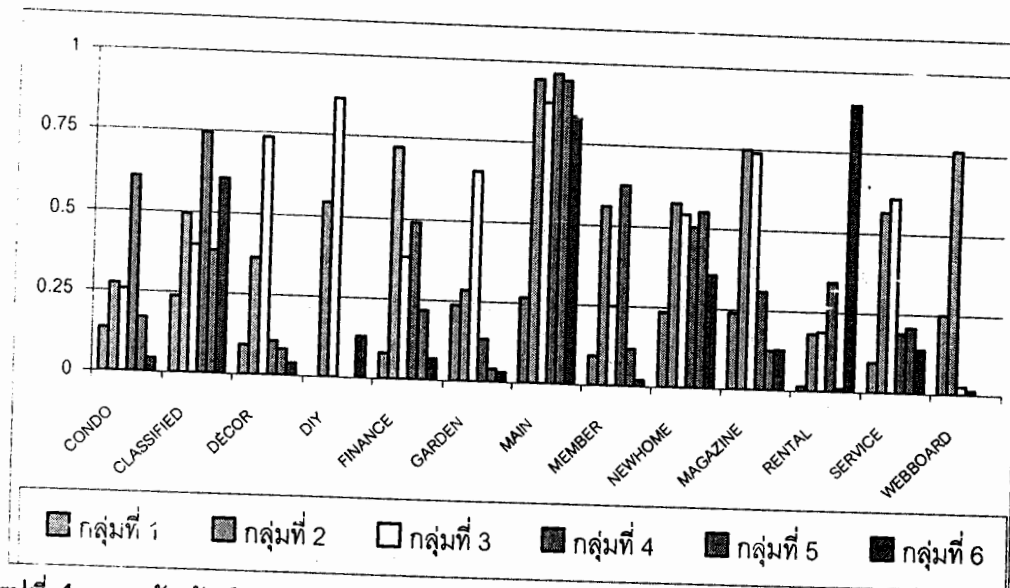
4.1 ผลการศึกษา

ผลจากการทำเหมืองข้อมูลด้วยเทคนิค Clustering

หลังการประมวลผลข้อมูลด้วยเทคนิค Clustering ข้อมูล Web Access Log 10,585 รายการได้ถูกแบ่งออกเป็น 6 กลุ่มที่มีขนาดแตกต่างกัน โดยกลุ่มที่ 5 มีขนาดใหญ่ที่สุด 5,300 รายการ, อันดับรองลงมา คือ กลุ่มที่ 4 มี 2,007 รายการ และกลุ่มที่ 2 มีขนาดเล็กที่สุด 239 รายการ โดยขนาดของกลุ่มข้อมูลทั้งหมดแสดงได้ดังรูปที่ 3 และความสัมพันธ์ระหว่างกลุ่มข้อมูลทั้งหมดกับประเภทของ Web Page แสดงได้ดังรูปที่ 4



รูปที่ 3 ขนาดของกลุ่มข้อมูลที่เป็นผลลัพธ์จากการใช้เทคนิค Clustering



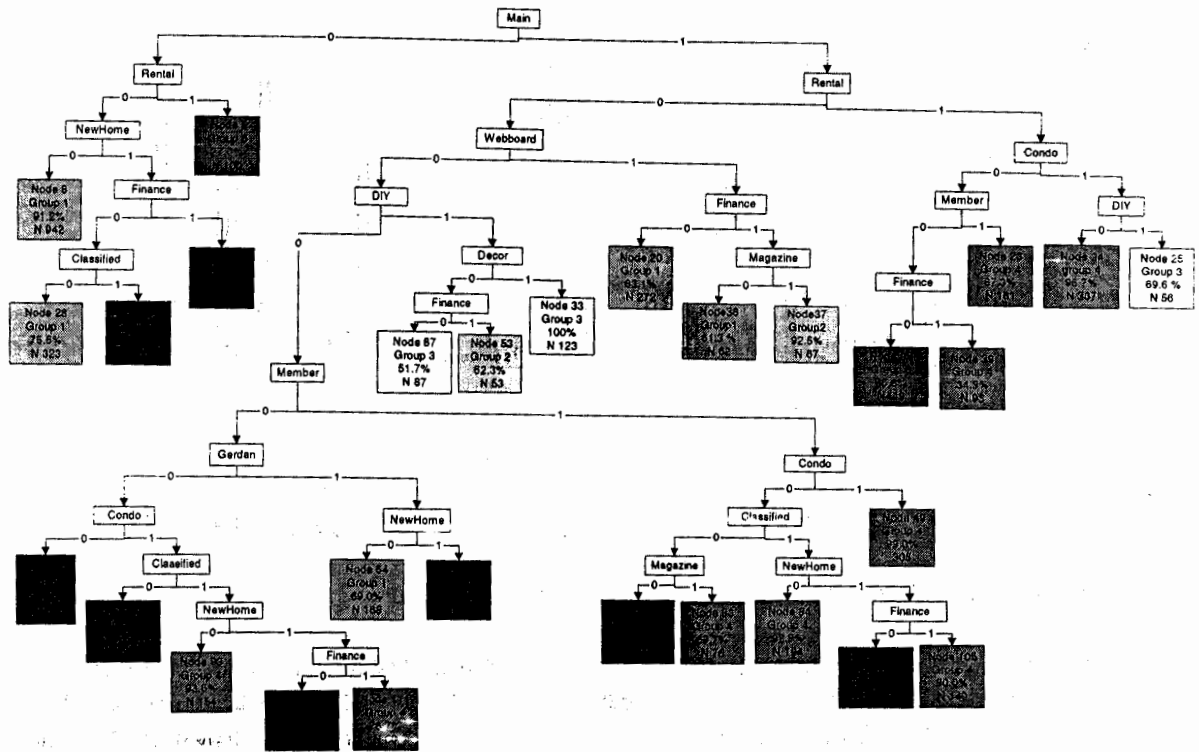
รูปที่ 4 ความสัมพันธ์ระหว่างกลุ่มข้อมูลกับประเภทของหน้าเว็บ

รูปที่ 4 แสดงความสัมพันธ์ระหว่างกลุ่มข้อมูลกับประเภทของหน้าเว็บ โดยแกน X แสดงกลุ่มของหน้าเว็บทั้ง 13 กลุ่มและ แกน Y แสดงระดับความสำคัญของกลุ่มของหน้าเว็บ (ตัวแปร) ที่มีต่อกลุ่มข้อมูลต่าง ๆ โดยค่าของแกน Y อยู่ระหว่าง 0 และ 1 ค่าที่เข้าใกล้ 1 แสดงให้ทราบว่าตัวแปรนั้นมีความสัมพันธ์กับกลุ่มข้อมูลนั้นสูง [8] จากรูปที่ 4 สามารถสรุปความสัมพันธ์ของกลุ่มข้อมูลกับประเภทของหน้าเว็บได้ดังนี้

- ตัวแปร Condo มีความสัมพันธ์กับกลุ่มที่ 4 มากที่สุด
- ตัวแปร Classified มีความสัมพันธ์กับกลุ่มที่ 4 มากที่สุดและมีความสัมพันธ์กับกลุ่มที่ 6 และกลุ่มที่ 2 เป็นอันดับรองลงมา
- ตัวแปร Décor มีความสัมพันธ์กับกลุ่มที่ 3 มากที่สุด
- ตัวแปร DIY มีความสัมพันธ์กับกลุ่มเพียง 2 กลุ่มคือ กลุ่มที่ 3 และกลุ่มที่ 2 โดยกลุ่มที่ 3 มีความสัมพันธ์กับตัวแปร DIY มากกว่ากลุ่มที่ 2
- ตัวแปร Finance มีความสัมพันธ์กับกลุ่มที่ 2 มากที่สุด
- ตัวแปร Garden มีความสัมพันธ์กับกลุ่มที่ 3 มากที่สุด
- ตัวแปร Main มีความสัมพันธ์สูงกับทุกกลุ่มยกเว้นกลุ่มที่ 1
- ตัวแปร Member มีความสัมพันธ์กับกลุ่มที่ 4 และกลุ่มที่ 2 มากที่สุดตามลำดับ
- ตัวแปร NewHome มีความสัมพันธ์กับกลุ่มที่ 2 กลุ่มที่ 5 กลุ่มที่ 3 และกลุ่มที่ 4 ตามลำดับ
- ตัวแปร Magazine มีความสัมพันธ์สูงกับกลุ่มที่ 2 และกลุ่มที่ 3
- ตัวแปร Rental มีความสัมพันธ์กับกลุ่มที่ 6 มากที่สุด
- ตัวแปร Service มีความสัมพันธ์สูงกับกลุ่มที่ 3 และกลุ่มที่ 2
- ตัวแปร Webboard มีความสัมพันธ์กับกลุ่มที่ 2 มากที่สุด

ผลจากการทำเหมืองข้อมูลด้วยเทคนิค Decision Tree

ผู้ศึกษาได้แบ่งข้อมูล 10,585 รายการออกเป็น 2 ส่วน คือ ส่วนที่ 1 ข้อมูลที่ใช้ในการสร้างแผนภาพต้นไม้ (Training data) ขนาด 8,448 รายการ และข้อมูลที่ใช้ทดสอบความถูกต้องของแผนภาพต้นไม้ (Test data) ขนาด 2,137 รายการ โดยผลลัพธ์ที่ได้จากเทคนิค Decision Tree เป็นแผนภาพต้นไม้ที่มี Leaves Node ทั้งหมด 29 Nodes ดังแสดงในรูปที่ 5 และมีความถูกต้อง 90.24% บนข้อมูลที่ใช้ทดสอบความถูกต้อง



รูปที่ 5 แผนภาพต้นไม้ที่ได้จากเทคนิค Decision Tree ตารางที่ 4 แสดงข้อมูลสรุปของ Node ในแผนภาพต้นไม้ที่มีความถูกต้องในการจำแนกข้อมูลสูงกว่า 70% โดยจำแนกตามกลุ่มของ Node เหล่านั้น ข้อมูลที่แสดงในตารางที่ 4 สามารถจำแนกออกเป็น 2 ส่วน คือ

กลุ่ม	ขนาดกลุ่ม	Node ID	ขนาดข้อมูล	Accuracy (%)	สัดส่วน	รวม	Main	Rental	Webboard	DIY	Member	Garden	Condo	New Home	Classified	Finance	Service	Magazine	Decor		
1	1767	8	942	91.2	53%	96%	0	0							0						
		28	323	75.5	18%		0	0							1	0	0				
		20	272	83.1	15%		1	0	1									0			
		64	168	69.0	10%		1	0	0	0	0	0	1		0						
2	120	37	67	92.5	56%	56%	1	0	1								1		1		
		33	123	100.0	38%		1	0	0	1											1
4	1464	25	56	69.6	17%	94%	1	1		1				1							
		24	337	96.7	23%		1	1		0				1							
		49	305	99.0	21%		1	0	0	0	1		1								
		84	194	92.8	13%		1	0	0	0	1		0	0	1						
		23	181	87.3	12%		1	1			1				0						
		93	154	93.5	11%		1	0	0	0	0	0	0	1	0	1					
		105	140	80.0	10%		1	0	0	0	1		0	1	1	1	1				
5	4226	111	60	76.7	4%	97%	1	0	0	0	0	0	0								
		63	2698	99.8	64%		1	0	0	0	0	0	0	0							
		75	547	95.2	13%		1	0	0	0	0	0	0	1		0					
		82	340	95.9	8%		1	0	0	0	1		0							0	
		65	198	72.7	5%		1	0	0	0	1		0								
		110	153	88.2	4%		1	0	0	0	0	0	1		1	1	1	0			
		29	85	78.8	2%		0	0							1	1	0				
6	549	104	88	80.7	2%	100%	1	0	0	0	1		0	1	1	0					
		38	449	86.6	82%		1	1			0					0					
		5	100	77.0	18%		0	1													

ตารางที่ 4 ข้อมูลสรุปของแผนภาพต้นไม้

- ส่วนที่ 1 ข้อมูลรายละเอียดของกลุ่มและ Node ในแผนภาพต้นไม้ที่อยู่ภายใต้กลุ่ม นั้น ๆ ตามลำดับดังนี้
 - กลุ่ม, ขนาดของข้อมูลในกลุ่ม, Node ID, ขนาดของข้อมูลที่อยู่ภายใต้ Node นั้น ๆ , ความแม่นยำในการจำแนกข้อมูลของ Node นั้น (Accuracy), สัดส่วนของข้อมูลที่อยู่ภายใต้ Node นั้นคิดเป็นสัดส่วนกับขนาดของข้อมูลในกลุ่ม, และเปอร์เซ็นต์รวมของข้อมูลในกลุ่มนั้นที่ Node ต่าง ๆ สามารถจำแนกได้ด้วย ความถูกต้องมากกว่าหรือเท่ากับ 70%
- ส่วนที่ 2 ข้อมูลแสดงลักษณะสำคัญของ Node ที่อยู่ภายใต้กลุ่มต่าง ๆ
 - ส่วนนี้ประกอบด้วยข้อมูล 13 คอลัมน์โดยแต่ละคอลัมน์คือกลุ่มของหน้าเว็บ 13 กลุ่ม โดยแต่ละช่องในตารางมีค่าที่เป็นไปได้ 3 ค่า คือ ค่าว่าง, ค่า 0 และ ค่า 1
 - ค่าว่าง หมายความว่า กรณีที่กลุ่มของหน้าเว็บนั้นไม่แสดงความสัมพันธ์ที่ชัดเจนกับ Node กลุ่มผู้เยี่ยมชมในแถว (Row) นั้น
 - ค่า 1 หมายความว่า กรณีที่กลุ่มของหน้าเว็บนั้นเป็นกลุ่มของหน้าเว็บที่ผู้เยี่ยมชมเว็บไซต์ที่จัดอยู่ใน Node นั้นเข้าเยี่ยมชม
 - ค่า 0 หมายความว่า กรณีที่กลุ่มของหน้าเว็บนั้นเป็นกลุ่มของหน้าเว็บที่ผู้เยี่ยมชมเว็บไซต์ที่จัดอยู่ใน Node นั้นเข้าไม่เข้าเยี่ยมชม

4.2 วิเคราะห์ผลการทำเหมืองข้อมูล

จากรูปที่ 4 และตารางที่ 4 พบว่า ผลที่ได้จากการทำ Clustering และ การใช้เทคนิค Decision Tree มีความสอดคล้องกัน โดยแสดงลักษณะสำคัญของผู้เข้าชมเว็บไซต์ HomeDD.com ดังนี้

กลุ่มที่ 1 สนใจเกี่ยวกับการซื้อบ้านมือสอง

กลุ่มนี้เป็นกลุ่มขนาดกลาง (1,767 รายการ คิดเป็น 18.27% ของข้อมูลทั้งหมด) ที่สนใจในเรื่องต่างๆ แตกต่างกัน ลักษณะรวมของผู้ใช้บริการกลุ่มนี้ คือ ไม่แสดงความสนใจในการเช่าที่อยู่อาศัย (Rental) ผู้ใช้บริการ 71% เข้าชมเว็บไซต์โดยไม่ผ่านหน้า Main และมีผู้ใช้บริการ 63% ไม่แสดงความสนใจในกลุ่มหน้าเว็บเกี่ยวกับบ้านใหม่ (New Home)

กลุ่มที่ 2 สนใจการขอสินเชื่อและเว็บบอร์ด

กลุ่มที่ 2 เป็นกลุ่มขนาดเล็ก (120 รายการ คิดเป็น 1.42% ของข้อมูลทั้งหมด) จากตารางที่ 4 พบว่า แผนภาพต้นไม้ที่สร้างขึ้นมีความแม่นยำในการจำแนกกลุ่มผู้ให้บริการกลุ่มนี้ไม่สูง (56%) อย่างไรก็ตามเมื่อเทียบกับการคาดเดากลุ่ม 1 ใน 6 กลุ่มอย่างสุ่มซึ่งมีความน่าจะเป็นของการทำนายได้ถูกต้อง 16.67% ความแม่นยำในการจำแนกกลุ่มของแผนภาพต้นไม้ในรูปที่ 5 นั้นมีความแม่นยำสูงกว่า 2.36 เท่า ลักษณะที่สำคัญของ Node ในแผนภาพต้นไม้ที่จัดอยู่ในกลุ่มนี้ คือ ผู้ใช้บริการมักเข้าใช้บริการผ่านหน้าหลัก มีความสนใจในหน้าเว็บกลุ่มที่เกี่ยวกับการขอสินเชื่อ (Finance) เข้าใช้บริการเว็บบอร์ด (Webboard) และไม่สนใจในการเช่าที่อยู่อาศัย (Rental)

เมื่อทำการศึกษาข้อความจากเว็บบอร์ด แล้วพบว่า ลักษณะดังกล่าวมีความสอดคล้องกัน นั่นคือ หัวข้อด้านการเงินเป็นหัวข้อที่ได้รับการ post เป็นส่วนใหญ่ อาจเพราะส่วนหนึ่งต้องการความรู้ในการขอสินเชื่อในรูปแบบการประมูลสินเชื่อ (Auction Loan) ซึ่งเป็นบริการที่ HomeDD.com ให้บริการ นอกจากนี้จากลักษณะสำคัญประการหนึ่งอีกของผู้ใช้บริการกลุ่มนี้ คือไม่สนใจในการเช่าที่อยู่อาศัย (Rental) ซึ่งสอดคล้องกับลักษณะสินเชื่อที่ HomeDD.com ให้บริการประมูล ซึ่งเป็นสินเชื่อสำหรับการซื้อที่อยู่อาศัยแต่ไม่ครอบคลุมสินเชื่อสำหรับการเช่าที่อยู่อาศัย

กลุ่มที่ 3 สนใจการซ่อมแซมด้วยตนเอง และการตกแต่งบ้าน

กลุ่มที่ 3 เป็นกลุ่มขนาดเล็ก (322 รายการ คิดเป็น 3.81% ของข้อมูลทั้งหมด) และแผนภาพต้นไม้ที่สร้างขึ้นมีความแม่นยำในการจำแนกกลุ่มผู้ใช้บริการกลุ่มนี้ไม่สูง (56%) ผู้ใช้บริการในกลุ่มนี้มีพฤติกรรมกระเเย่มชมผ่านทางหน้า Main และสนใจในเรื่องของการซ่อมแซมด้วยตนเอง (DIY) 38% ของผู้ใช้บริการสนใจในหน้าเว็บในกลุ่มของการออกแบบตกแต่งบ้าน (Décor) แต่ไม่สนใจในการเช่าที่อยู่อาศัย (Rental) และไม่สนใจการเช่าที่อยู่อาศัย (Rental)

กล่าวได้ว่ากลุ่มที่ 3 เป็นกลุ่มขนาดเล็กที่สนใจในเรื่องของการซ่อมแซมด้วยตนเอง (DIY) และส่วนใหญ่สนใจการตกแต่งบ้าน (Décor) ซึ่งสอดคล้องกันเพราะเนื้อหาของหน้าเว็บในทั้ง 2 กลุ่มนี้สอดคล้องกัน แต่ไม่สนใจการเช่าที่อยู่อาศัย เพราะคาดว่าเป็นกลุ่มผู้ใช้บริการที่มีที่อยู่อาศัยอยู่แล้วและต้องการเพียงตกแต่งหรือซ่อมแซมที่อยู่อาศัย

กลุ่มที่ 4 สนใจซื้อหรือขายคอนโดมิเนียม

กลุ่มที่ 4 เป็นกลุ่มผู้ใช้บริการขนาดกลาง (1,464 รายการ คิดเป็น 17.33%) ที่มีพฤติกรรมการใช้บริการผ่านหน้า Main ทั้งหมด ลักษณะสำคัญของผู้ใช้บริการกลุ่มนี้ คือ 82% ไม่สนใจในการซ่อมแซมที่อยู่อาศัยด้วยตัวเอง (DIY) 59% ไม่ใช้บริการเว็บบอร์ด 59% ไม่สนใจการเช่าซื้อที่อยู่อาศัย 59% เข้าชมหน้าเว็บเกี่ยวกับคอนโดมิเนียม 56% เป็นสมาชิกของเว็บไซต์ และ 38% เข้าชมหน้าเว็บเกี่ยวกับการซื้อขายบ้าน (Classified)

กล่าวได้ว่าผู้ใช้บริการในกลุ่มนี้แสดงพฤติกรรมของผู้ที่สนใจซื้อหรือขายคอนโดมิเนียม โดยมีลักษณะของความเป็นสมาชิกเข้ามาเกี่ยวข้อง เนื่องจากด้วย HomeDD.com มีข้อบังคับให้ ผู้ที่สามารถประกาศขายอสังหาริมทรัพย์ได้ ต้องเป็นสมาชิกของเว็บไซต์เท่านั้น

กลุ่มที่ 5 สนใจซื้อบ้าน

กลุ่มที่ 5 เป็นกลุ่มขนาดใหญ่ที่สุด (4,226 รายการ คิดเป็น 50.02%) 95% มีพฤติกรรมกระเเย่มชมผ่านทางหน้า Main ลักษณะสำคัญที่มีร่วมกันคือ 97% ไม่มีความสนใจในการเช่าที่อยู่อาศัย (Rental) 95% ไม่สนใจบริการเว็บบอร์ด (Webboard) และไม่สนใจการตกแต่งซ่อมแซมบ้านด้วยตนเอง (DIY) 86% ไม่ได้เป็นสมาชิกของเว็บไซต์ (Member) 74% ไม่สนใจในหน้าเว็บที่เกี่ยวกับคอนโดมิเนียม (Condo) 13% สนใจข้อมูลเกี่ยวกับบ้านใหม่ (New Home) และ 8% สนใจเกี่ยวกับการประกาศซื้อขายบ้าน (Classified)

กล่าวได้ว่าผู้ใช้บริการในกลุ่มนี้เป็นผู้สนใจซื้อบ้าน เนื่องจากไม่ได้แสดงความสนใจในการเช่าที่อยู่อาศัย และส่วนใหญ่ไม่แสดงความสนใจเกี่ยวกับคอนโดมิเนียม ผู้ใช้ส่วนใหญ่ไม่เข้าใช้บริการหน้าของสมาชิก ซึ่งสะท้อนให้เห็นว่า ผู้ใช้บริการกลุ่มนี้ไม่มีความประสงค์ในการขายบ้าน

กลุ่มที่ 6 สนใจเช่าที่อยู่อาศัย

กลุ่มที่ 6 เป็นกลุ่มขนาดเล็ก (549 รายการ คิดเป็น 6.50%) ที่มีพฤติกรรมสนใจในเรื่องการเช่าที่อยู่อาศัย (Rental) เป็นหลัก 82% ไม่เข้าใช้บริการหน้าของสมาชิก (Member) ไม่สนใจคอนโดมิเนียม (Condo) และไม่สนใจบริการสินเชื่อ (Finance)

กล่าวได้ว่า ผู้ใช้บริการกลุ่มนี้สนใจเช่าที่อยู่อาศัยเพียงอย่างเดียว จึงไม่สนใจเกี่ยวกับหน้าเว็บที่เกี่ยวกับคอนโดมิเนียมซึ่งเป็นการซื้อหรือขายคอนโดมิเนียม อีกทั้งไม่มีความจำเป็นต้องเป็นสมาชิกของเว็บไซต์ และไม่มีความจำเป็นต้องขอสินเชื่อ

4.3 การประยุกต์ใช้

1. การจัดทำชุมชนบนเว็บไซต์

ผลการศึกษาแสดงให้เห็นว่าผู้ใช้บริการเว็บไซต์ HomeDD.com สามารถแบ่งเป็นกลุ่มตามความสนใจได้ ดังนั้นการสร้างชุมชนให้เกิดขึ้นในเว็บไซต์เป็นกลยุทธ์หนึ่งที่สำคัญในการบริหารความสัมพันธ์กับผู้ให้บริการเว็บไซต์ [1] โดยอาจสร้างชุมชนผ่านเว็บบอร์ด อย่างไรก็ตาม เมื่อพิจารณาเว็บบอร์ดของ HomeDD.com ในช่วงที่เก็บข้อมูลนั้น พบว่ามีการจัดไว้ 4 หมวดหมู่ คือ หมวดตกแต่งบ้าน หมวดซ่อมแซม หมวดจัดสวน และ หมวดรำพึงรำพันเรื่องบ้าน ซึ่งไม่สอดคล้องกับกลุ่มที่ศึกษาได้ ดังนั้น HomeDD.com อาจทำการปรับหมวดหมู่ในเว็บบอร์ดเพื่อตอบสนองความต้องการของผู้ใช้บริการดังนี้

- กลุ่มบ้านใหม่ เป็นชุมชนสำหรับผู้ให้บริการที่มีความต้องการในการซื้อบ้านใหม่ในโครงการต่างๆ ทั้งยังเป็นช่องทางการประชาสัมพันธ์ที่ดีให้กับโครงการบ้านใหม่ ซึ่งเป็นธุรกิจหลักของบริษัทแลนด์แอนด์เฮาส์ซึ่งเป็นเจ้าของเว็บไซต์ HomeDD.com
- กลุ่มคอนโดมิเนียม เป็นชุมชนสำหรับผู้สนใจในคอนโดมิเนียม ให้มีศูนย์กลางในการแลกเปลี่ยนข้อมูล นอกเหนือจากข้อมูลที่มีในเว็บไซต์
- กลุ่มบ้านมือสอง เป็นชุมชนสำหรับผู้สนใจซื้อขายบ้าน ให้สามารถแลกเปลี่ยนข้อมูลกันเพิ่มเติมได้ นอกเหนือจากการลงประกาศซื้อขายเพียงอย่างเดียว
- กลุ่มเช่าที่อยู่อาศัย เป็นชุมชนทางเว็บไซต์สำหรับสื่อสารระหว่างผู้สนใจให้เช่าและผู้สนใจเช่าบ้าน
- กลุ่มการขอสินเชื่อ เป็นชุมชนสำหรับผู้สนใจขอสินเชื่อ ทั้งยังเป็นช่องทางในการแนะนำบริการการประมูลสินเชื่อ (Auction Loan) ซึ่งเป็นบริการพิเศษของ HomeDD.com ด้วย
- กลุ่มซ่อมแซมและตกแต่ง เป็นชุมชนสำหรับผู้ที่มีที่อยู่อาศัยแล้วต้องการตกแต่งหรือซ่อมแซมที่อยู่อาศัย ให้สามารถแลกเปลี่ยนข้อมูลการตกแต่งหรือการซ่อมแซมได้

2. เพิ่มบริการให้คำปรึกษาการขอสินเชื่อ

ผลการศึกษาแสดงให้เห็นว่ากลุ่มที่สนใจการขอสินเชื่อ มักจะสนใจการใช้บริการเว็บบอร์ดควบคู่กัน ทั้งนี้อาจเป็นเพราะผู้ที่ขอสินเชื่อต้องการศึกษาข้อมูลและขอคำแนะนำเพิ่มเติมในการขอสินเชื่อ ดังนั้นจึงควรเพิ่มบริการให้คำปรึกษาแก่ผู้สนใจขอสินเชื่อ โดยอาจจัดให้อยู่ในรูปแบบ Web Master ที่เป็นผู้ให้คำแนะนำเกี่ยวกับการขอสินเชื่อโดยตรง

3. รวมกลุ่มหน้าเว็บเรื่องการตกแต่งบ้าน (Décor) และการซ่อมแซมด้วยตนเอง (DIY)

จากการศึกษา พบว่ากลุ่มที่สนใจการซ่อมแซมด้วยตนเองจะสนใจการตกแต่งบ้านควบคู่กันด้วย ดังนั้นจึงสามารถที่จะรวมกลุ่มของหน้าเว็บทั้งสองกลุ่มไว้ด้วยกัน โดยให้การซ่อมแซมด้วยตนเองเป็นกลุ่มย่อยของการตกแต่งบ้าน เพื่อให้ผู้ใช้บริการสามารถเข้าถึงข้อมูลที่ต้องการได้สะดวกสบายมากยิ่งขึ้น โดยไม่จำเป็นต้อง Navigate ข้ามกลุ่มไปมา

5. ข้อจำกัดของผลการศึกษา

การศึกษานี้เป็นการศึกษาบน Web Access Log ของเว็บไซต์ HomeDD.com ผลที่ได้จากการศึกษาและผลการวิเคราะห์ไม่สามารถนำไปประยุกต์ใช้ในการบริหารเว็บไซต์อื่นได้โดยตรง นอกจากนี้ ผลที่ได้จากการศึกษาเป็นพฤติกรรมของผู้ใช้บริการในช่วงที่เว็บไซต์ HomeDD.com มีโครงสร้างเหมือนกับในช่วงที่เก็บข้อมูลเพื่อทำการศึกษาเท่านั้น

6. แนวทางการศึกษาต่อเนื่อง

การศึกษาค้างนี้ เน้นไปที่ประเภทของหน้าเว็บที่ผู้ใช้บริการเข้าเยี่ยมชมในช่วง (Session) ของ การเข้าชมเว็บไซต์ครั้งหนึ่ง ๆ จำนวนหน้ารวมทั้งลำดับของหน้าที่ผู้ใช้บริการเข้าเยี่ยมชมหน้าเว็บ ในแต่ละกลุ่มไม่ได้ถูกนำมาพิจารณาในการศึกษา การดำเนินการศึกษาต่อเนื่องในเรื่องดังกล่าว อาจนำมาสู่การประยุกต์ใช้ผลในด้านการปรับปรุงโครงสร้างของหน้าหน้าเว็บให้สอดคล้องกับ ความต้องการของผู้ใช้บริการมากยิ่งขึ้น เช่น มีการจัดลำดับหน้าที่มีผู้ใช้บริการมากให้อยู่ใน ลำดับขั้นที่ไม่ไกลจากหน้าหลัก การจัดหน้าเว็บที่ผู้ใช้บริการมักเรียกดูความถี่กันให้มีลิงค์เชื่อม โยงระหว่างกัน หรือ การรวมหน้าเว็บที่มีลำดับการเรียกดูต่อเนื่องกันไว้ด้วยกัน หรือสามารถเชื่อม โยงระหว่างกันได้โดยสะดวก เป็นต้น

7. บทสรุป

การศึกษานี้ได้แสดงให้เห็นถึงวิธีการในการสร้างประโยชน์ในทางอ้อมจากพาณิชย์ อิเล็กทรอนิกส์ นั่นคือการทำเหมืองข้อมูลเว็บไซต์ หรือ การประยุกต์ใช้เหมืองข้อมูลกับข้อมูลใน Web Access Log จากกรณีศึกษาพบว่า การทำเหมืองข้อมูลเว็บไซต์ได้ทำให้องค์กรทราบถึง สารสนเทศใหม่ ๆ เกี่ยวกับผู้ใช้บริการเว็บไซต์หรือผู้บริโภคของพาณิชย์อิเล็กทรอนิกส์ เช่น กลุ่ม ของผู้ใช้บริการที่ได้มีการจัดแบ่งตามพฤติกรรมการใช้งานเว็บไซต์จริง สารสนเทศเหล่านี้ สามารถนำไปประยุกต์ใช้ในการบริหารจัดการเว็บไซต์ รวมถึงการบริการจัดการองค์กรให้ตอบ สนองความต้องการของผู้บริโภคได้ดีมากยิ่งขึ้น ซึ่งถือได้ว่าเป็นอีกแนวทางหนึ่งในการเติมเต็ม ข้อต่อของพาณิชย์อิเล็กทรอนิกส์ในการที่ไม่สามารถมีปฏิสัมพันธ์กับผู้บริโภคได้โดยตรง หรือ กล่าวได้ว่า การทำเหมืองข้อมูลเว็บไซต์สามารถใช้เป็นวิธีหนึ่งในการบริหารความสัมพันธ์กับลูกค้า (CRM) ผ่านเครือข่ายอิเล็กทรอนิกส์

8. กิจกรรมประกาศ

ผู้ศึกษาขอขอบคุณบริษัทแลนด์แอนด์เฮาส์ จำกัด (มหาชน) ที่ให้ความอนุเคราะห์ข้อมูลสำหรับ ใช้เป็นกรณีศึกษา พร้อมทั้งให้รายละเอียดข้อมูลเกี่ยวกับเว็บไซต์ ซึ่งสามารถนำมาใช้ในการ ศึกษาได้เป็นอย่างดี

9. เอกสารอ้างอิง

- [1] ณัฐพงษ์ พันธุ์เกียรติไพศาล และ ดาว ไวรักษ์สัตว์. *กลยุทธ์ทางการตลาดให้กับ เว็บไซต์*. บริษัทซีเอ็ดยูเคชั่น จำกัด, กรุงเทพฯ, 2544.
- [2] <http://www.apache.org>
- [3] <http://www.homedd.com>
- [4] <http://www.webtrends.com>
- [5] <http://www.sas.com>
- [6] R. Kosala & H. Blockeel, Web Mining Research: A Survey. *SIGKDD Explorations*, 2(1):1-15, Jun 2000.
- [7] J. Srivastava et al., Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 1(2):12-23, Jan 2000.
- [8] S. Walsh et al., *Applying Data Mining Techniques Using Enterprise Miner Course Notes*. SAS Institute Inc., USA, 2002.