Review Paper:

ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนคิการทำเหมืองข้อมูล A Prediction System for Undergraduate Student Dropout at Faculty of Science, Buriram Rajabhat University using Data Mining Techniques

นนทวัฒน์ ทวีชาต และคณะ (2564) ทำการวิจัยเพื่อ 1) เพื่อสร้างแบบจำลองการทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล 2) เพื่อพัฒนาระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภฏั บุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้คลังข้อมูล จากฐานข้อมูลการศึกษา ที่จัดเก็บข้อมูลของนักศึกษาตั้งแต่ปีการศึกษา 2556 - 2559 โดยใช้ข้อมูล ตัวอย่างจากนักศึกษาคณะวิทยาศาสตร์ ด้วยมีอัตราการพ้นสภาพมากที่สุดในมหาวิทยาลัยราชภัฏ บุรีรัมย์ จากนั้นนำข้อมูลจากหลายตารางมารวมกันผ่านกระบวนการเตรียมข้อมูลทำให้ได้ข้อมูลจำนวน 3,604 ระเบียน โดยใช้ อัลกอริทึมสำหรับการจำแนกประเภทเป็น J48 และใช้การวัดประสิทธิภาพการจำแนกประเภทโดยวิธีการตรวจสอบไขว้ (K-fold cross-validation) และนำ Model ที่ได้ไปปรับใช้กับ Web application โดยเลือกใช้ ภาษา PHP HTML JavaScript และระบบจัดการฐานข้อมูล MySQL โดยพบว่า 1) การใช้งานข้อมูลที่เก็บไว้ในฐานข้อมูลเพื่อการศึกษาของมหาวิทยาลัยราชภัฏบุรีรมัย์มาทำการวิเคราะห์ที่มีตารางจำนวนมาก และข้อมูลมีหลากหลายรูปแบบ รวมถึงการมีค่าผิดพลาด (Missing value) เก็บไว้จำนวนมาก ดังนั้นในขั้นตอนการเตรียมข้อมูลจึงต้องวางแผนเป็นอย่างดี ซึ่งต้องใช้เวลานาน และต้องกระทำอย่างระมัดระวัง 2) ความเหมาะสมของแอททริบิวส์ สำหรับการจัดประเภทข้อมูล (Data classification) ค่าข้อมูลจะต้องมีค่าที่วนซ้ำ และไม่หลากหลาย 3) อัตราการพ้นสภาพของนกัศึกษาชั้นปีที่ 1 มีอตัราที่สูงกว่าช้ัน ปีอื่น ดังนั้นมหาวิทยาลัยจะต้องให้ความสำคัญสำหรับนักศึกษาใหม่ทั้งในเรื่องของผลการเรียนและ พฤติกรรมของผู้เรียนด้วย

วลัยนุช สกุลนุ้ย (2554) ทำการวิจัยเพื่อ 1) เพื่อพัฒนาระบบแนะนําหนังสือคอมพิวเตอร์แบบออนไลน์โดยใช้เทคนิคการกรองแบบอิง เนื้อหามาใช้กับการแนะนําหนังสือให้ตรงกับความต้องการของผู้ใช้งานโดยอัตโนมัติได้ 2) เพื่อประเมินความพึงพอใจของการพัฒนาระบบแนะนําหนังสือคอมพิวเตอร์แบบออนไลน์ที่ พัฒนาขึ้น โดยพัฒนาระบบที่มีการลงทะเบียนผู้ใช้งานหลังจากนั้นระบบจะทำการให้ผู้ใช้งานให้คะแนนกับหนังสือต่างๆเพื่อนำคะแนนเหล่านั้นไปเป็นตัวชี้วัดในระบบการแนะนำโดยผ่านเทคนิค Content-Based Filtering เมื่อผู้ใช้เข้ามาใช้ระบบครั้งต่อไปตัวระบบจะทำการแนะนำหนังสือตามความต้องการของผู้ใช้งานโดยอัตโนมัติ โดยใช้อัลกอริทึม Naïve Bayes Theorem โดยจะนำคะแนนความชอบของผู้ใช้งานต่อหนังสือ, ผู้แต่ง, เนื้อหาของหนังสือ มาเป็น metric ในการทำระบบแนะนำพบว่า 1) การแสดงรายละเอียดหนังสือ การค้นหาข้อมูลหนังสือ การ แนะนําหนังสือโดยอัตโนมัติ รายงานการให้คะแนนหนังสือของผู้ใช้แต่เรื่อง ยอมรับความพึงพอใจ การใช้งานอยู่ในระดับดี 2) การแนะนําหนังสือให้ตรงกับความ ต้องการของผู้ใช้โดยอัตโนมัติ ยอมรับความพึงพอใจการใช้งานอยู่ในระดับดี 3) ความเหมาะสมของปริมาณข้อมูลที่นําเสนอในแต่ละหน้าจอการแสดงผลข้อมูลมีความเป็นรูปแบบและเป็ นมาตรฐานเดียวกันยอมรับความพึงพอใจการใช้งานอยู่ในระดับดี 4) ในการเข้าใช้งานการเตือนเมื่อพบข้อผิดพลาดในกรณีที่ ผู้ใช้ไม่ป้อนข้อมูลตามที่กําหนดความเหมาะสมของระบบในการรักษาความปลอดภัยยอมรับความ พึงพอใจการใช้งานอยู่ในระดับดี

**เอกสารอ้างอิง**

นนทวัฒน์ ทวีชาต และคณะ (2564)**.** ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนคิการทำเหมืองข้อมูล

วลัยนุช สกุลนุ้ย (2554)**.** วิเคราะห์และพัฒนาระบบแนะนําหนังสือคอมพิวเตอร์ แบบออนไลน์โดยใช้เทคนิคการกรองแบบอิงเนื้อหา

**เรื่อง :** ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนคิการทำเหมืองข้อมูล

**วัตถุประสงค์ :** 1) เพื่อสร้างแบบจำลองการทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล 2) เพื่อพัฒนาระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภฏั บุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล

**ขั้นตอนการดำเนินงาน :** ใช้คลังข้อมูล จากฐานข้อมูลการศึกษา ที่จัดเก็บข้อมูลของนักศึกษาตั้งแต่ปีการศึกษา 2556 - 2559 โดยใช้ข้อมูล ตัวอย่างจากนักศึกษาคณะวิทยาศาสตร์ ด้วยมีอัตราการพ้นสภาพมากที่สุดในมหาวิทยาลัยราชภัฏ บุรีรัมย์ จากนั้นนำข้อมูลจากหลายตารางมารวมกันผ่านกระบวนการเตรียมข้อมูลทำให้ได้ข้อมูลจำนวน 3,604 ระเบียน โดยใช้ อัลกอริทึมสำหรับการจำแนกประเภทเป็น J48 และใช้การวัดประสิทธิภาพการจำแนกประเภทโดยวิธีการตรวจสอบไขว้ (K-fold cross-validation) และนำ Model ที่ได้ไปปรับใช้กับ Web application โดยเลือกใช้ ภาษา PHP HTML JavaScript และระบบจัดการฐานข้อมูล MySQL

**การดำเนินงาน**

1. วางกรอบการศึกษา
   1. การเตรียมข้อมูลก่อนประมวลผล (Data pre-processing)
   2. อัลกอริทึมสำหรับการจำแนกประเภท (Classifier algorithm)
   3. การวัดประสิทธิภาพการจำแนกประเภท (Evaluation classifier)
   4. การสร้างแบบจำลอง (Academic DSS model)
   5. การพฒันาและประเมินผลระบบ (Development and evaluation system)
2. การรวบรวมข้อมูลเพื่อการประมวลผล
3. การคัดเลือกและแปลงข้อมูล (Data selection and transformation)
4. การวิเคราะห์ข้อมูล
   1. การสร้างและทดสอบแบบจำลอง
   2. การวัดค่าประสิทธิภาพของตัวแบบการพยากรณ์
5. การพัฒนาและการประเมินผลระบบ

**ผลลัพท์ :** 1) การใช้งานข้อมูลที่เก็บไว้ในฐานข้อมูลเพื่อการศึกษาของมหาวิทยาลัยราชภัฏบุรีรมัย์มาทำการวิเคราะห์ที่มีตารางจำนวนมาก และข้อมูลมีหลากหลายรูปแบบ รวมถึงการมีค่าผิดพลาด (Missing value) เก็บไว้จำนวนมาก ดังนั้นในขั้นตอนการเตรียมข้อมูลจึงต้องวางแผนเป็นอย่างดี ซึ่งต้องใช้เวลานาน และต้องกระทำอย่างระมัดระวัง 2) ความเหมาะสมของแอททริบิวส์ สำหรับการจัดประเภทข้อมูล (Data classification) ค่าข้อมูลจะต้องมีค่าที่วนซ้ำ และไม่หลากหลาย 3) อัตราการพ้นสภาพของนกัศึกษาชั้นปีที่ 1 มีอตัราที่สูงกว่าช้ัน ปีอื่น ดังนั้นมหาวิทยาลัยจะต้องให้ความสำคัญสำหรับนักศึกษาใหม่ทั้งในเรื่องของผลการเรียนและ พฤติกรรมของผู้เรียนด้วย

**ปัญหา :** Attribute GPA เนื่องด้วยตัว GPA มีข้อมูลที่หลากหลายมากซึ่งไม่เหมาะกับอัลกอริทึมการจำแนกอย่างมาก

**วิธีการแก้ไข :** ปรับให้อยู่ในรูปที่จำแนกได้ง่ายโดยแบ่งข้อมูลออกเป็นช่วงๆ โดย Week = GPA< 1.6 Medium = GPA 1.6-1.99 Good = GPA 2.0-2.5 Best = GPA>2.5

**เรื่อง :** วิเคราะห์และพัฒนาระบบแนะนําหนังสือคอมพิวเตอร์ แบบออนไลน์โดยใช้เทคนิคการกรองแบบอิงเนื้อหา

**วัตถุประสงค์ :** 1) เพื่อพัฒนาระบบแนะนําหนังสือคอมพิวเตอร์แบบออนไลน์โดยใช้เทคนิคการกรองแบบอิง เนื้อหามาใช้กับการแนะนําหนังสือให้ตรงกับความต้องการของผู้ใช้งานโดยอัตโนมัติได้ 2) เพื่อประเมินความพึงพอใจของการพัฒนาระบบแนะนําหนังสือคอมพิวเตอร์แบบออนไลน์ที่พัฒนาขึ้น

**ขั้นตอนการดำเนินงาน และ เทคนิคหรือ theory ที่ใช้ :**

**ระบบให้การแนะนํา**

ระบบให้การแนะนํา.(Recommender System) เป็นเทคโนโลยี สมัยใหม่ที่ถูกนํามาใช้ในการแนะนําข้อมูลต่างๆ ที่คาดว่าผู้ใช้น่าจะสนใจ หรืออาจจะเป็นข้อมูลที่ ผู้ใช้ต้องการ และเหมาะสําหรับการดําเนินธุรกิจแบบพาณิชย์อิเล็กทรอนิกส์ (e-Commerce) โดยทั่วไประบบให้การแนะนําประกอบด้วย 3 ส่วน ดังต่อไปนี้

1. ส่วนข้อมูลพื้นฐานที่จําเป็นต้องใช้ในการประมวลผล เช่น โปรไฟล์ของผู้ใช้แต่ละคน
2. ส่วนการป้อนข้อมูล เป็นข้อมูลที่ได้จากการป้อนข้อมูลเข้ามาของผู้ใช้ เช่น การให้ คะแนนเรตติ้ง ซึ่งมีอยู่ 2 แบบ ดังนี้
   1. แบบชัดเจน.(Explicit) เรตติ้งแบบชัดเจนจะแสดงอยู่ในรูปของจํานวนตัวเลข ตามระดับความนิยมตั้งแต่ 1 ถึง 5, 1 ถึง 10 หรือระดับอื่นๆ ขึ้นอยู่กับการใช้งาน
   2. แบบไม่ชัดเจน (Implicit) ส่วนเรตติ้งแบบไม่ชัดเจนได้มาจากพฤติกรรมการใช้ งานของผู้ใช้ต่างๆ เช่น ประวัติการซื้อสินค้าหรือประวัติการเข้ามาใช้งานของผู้ใช้ในอดีตที่ผ่านมา
3. ส่วนอัลกอริธึมเป็นส่วนสําคัญที่สุดที่ใช้ประมวลข้อมูลเพื่อให้การแนะนําชิ้นข้อมูล ออกมา

Content-Based Filtering

เป็นวิธีที่นิยมใช้ในการค้นคืนข้อมูลที่ให้ความสนใจกับคุณภาพของเนื้อหาข้อมูลเป็นหลัก โดยจะสนใจว่าลักษณะข้อมูลนั้นตรงตามโปรไฟล์ของผู้ใช้หรือไม่ ซึ่งถ้าใช่ก็จะนําเสนอข้อมูลนั้น ทันที.แต่ถ้าไม่ใช่ก็จะไม่สนใจแม้ว่าข้อมูลนั้น จะมีลักษณะใกล้เคียงกับข้อมูลที่ผู้ใช้ ต้องการก็ตาม

โดยนําเนื้อหาของข้อมูล เช่น คําสําคัญ (Keywords), วลี (Phrases) หรือคุณลักษณะ (Feature) มา สร้างเป็นโปรไฟล์ของผู้ใช้แต่ละคน เพื่อค้นหาข้อมูลที่ผู้ใช้คนนั้นสนใจซึ่งวิธีการของ Content Based Filtering จะไม่ประสบกับปัญหาการให้เรตติ้งต่อชิ้นข้อมูลที่ไม่ทั่วถึง และปัญหาชิ้นข้อมูลที่ ยังไม่ได้ให้เรตติ้ง

**ขั้นตอนการทํางานของเทคนิคการกรองแบบอิงเนื้อหา**

3 ขั้นตอน คือ

1. การสร้างเมตริกซ์ของผู้ใช้ชิ้นข้อมูลจริง

นำข้อมูลจากฐานข้อมูลที่มีอยู่มาทำเป็น เมตริกซ์ผู้ใช้ชิ้นข้อมูล ซึ่งลักษณะเป็นตารางที่มีแถวเป็นผู้ใช้ และมีคอลัมน์เป็นชิ้นข้อมูลซึ่งก็คือ ชื่อเรื่องของหนังสือ และข้อมูลในแต่ละช่องเป็นเรตติ้ง ที่ผู้ใช้คนนั้นให้กับหนังสือเรื่องนั้นๆ โดยเรตติ้งที่ให้มีค่าเป็นจํานวนเต็มตั้งแต่ 1-5 และเป็น Null ถ้าผู้ใช้ไม่เคยให้เรตติ้งกับหนังสือเรื่องนั้น

1. การทํานายเรตติ้งด้วยเทคนิคการกรองแบบอิงเนื้อหา

นําข้อมูลพื้นฐานที่มี สองค่าด้วยกัน ได้แก่ คะแนนชื่อเรื่องและเนื้อหาหนังสือ , คะแนนชื่อผู้แต่งหนังสือ โดยทําการแบ่งค่าเรตติ้ง จาก 1-5 ออกเป็น 5 คลาส คือ ค่าเรตติ้ง 1-5 โดยเรียงลําดับคะแนนจากระดับน้อยซึ่งก็คือ 1 ไปจนถึง ระดับสูงซึ่งก็คือ 5 นําข้อมูลเนื้อหาของหนังสือในฐานข้อมูล และเมตริกซ์ผู้ใช้ชิ้นข้อมูลของแต่ละประเภทมาให้ Classifier เรียนรู้ประวัติการให้เรต ติ้งของผู้ใช้แล้วสร้างเป็นโปรไฟล์ของผู้ใช้แต่ละคน ซึ่งเก็บข้อมูลความน่าจะเป็นที่เนื้อหาต่างๆ ของชิ้นข้อมูลจะจัดอยู่ในคลาส ต่าง ๆ จากนั้นนําโปรไฟล์ของผู้ใช้แต่ละคนไปทํานายเรตติ้งให้กับชิ้นข้อมูลที่ผู้ใช้นั้นๆ ที่ยังไม่มีการให้เรตติ้งโดยใช้ Naïve Bayes Theorem

Naïve Bayes Theorem

การเรียนรู้แบบ Naïve Bayes เป็นอัลกอริธึมที่ง่ายและเร็วในการคํานวณ และเป็นการเรียนรู้ที่ อาศัยหลักทางสถิติและความน่าจะเป็น โดยผลลัพธ์ที่ให้ค่าความน่าจะเป็นสูงสุดเป็น คําตอบของ

Diagram

Description automatically generated with medium confidence

Text

Description automatically generated

1. การสร้างเมตริกซ์ผู้ใช้ชิ้นข้อมูลเทียม

จากเวคเตอร์ผู้ใช้ชิ้นข้อมูลเทียมของผู้ใช้ทีละคนที่ได้มาด้วยวิธีการข้างต้น จากนั้นนําเวคเตอร์เทียม

ที่ได้จากผู้ใช้ทั้งหมดนํามารวมกันเป็นเมตริกซ์ผู้ใช้ชิ้นข้อมูลเทียมที่ไม่มีความเบาบาง หรือก็ได้ทำ

การทำนายค่าเหล่านั้นผ่าน อัลกอริทึม Naïve Bayes

**Web Application**

**เครื่องมือที่ใช้ในการพัฒนา**

* HyperText Markup Language (HTML)
* Personal Home Page Tool (PHP)
* ระบบฐานข้อมูลด้วยมายเอสคิวแอล(MySql)

**การดำเนินงาน**

1. การศึกษาข้อมูลและปัญหาของระบบงานเดิม
   1. การศึกษาปัญหาของระบบงานเดิม
   2. การศึกษาเครื่องมือที่ทําการพัฒนาเว็บไซต์
2. การออกแบบระบบ
   1. แผนภาพกระแสข้อมูลระดับสูง(Context Diagram)
   2. แผนภาพการไหลของข้อมูล ระดับที่ 0 (Data Flow Diagram Level 0) ของระบบ
   3. แผนภาพแสดงความสัมพันธ์ (Entity Relationship Model)
   4. วิเคราะห์โครงสร้างฐานข้อมูลระบบ
   5. หน้าจอเปลี่ยนแปลงรหัสผ่านในการเข้าสู่ระบบ
   6. หน้าจอการลงคะแนนความชอบต่อหนังสือของผู้ใช้
   7. หน้าจอจัดการข้อมูลหนังสือ
3. การพัฒนาระบบ
4. การทดสอบระบบ
5. การประเมินผลระบบ

**ผลลัพท์ :** 1) การแสดงรายละเอียดหนังสือ การค้นหาข้อมูลหนังสือ การ แนะนําหนังสือโดยอัตโนมัติ รายงานการให้คะแนนหนังสือของผู้ใช้แต่เรื่อง ยอมรับความพึงพอใจ การใช้งานอยู่ในระดับดี 2) การแนะนําหนังสือให้ตรงกับความ ต้องการของผู้ใช้โดยอัตโนมัติ ยอมรับความพึงพอใจการใช้งานอยู่ในระดับดี 3) ความเหมาะสมของปริมาณข้อมูลที่นําเสนอในแต่ละหน้าจอการแสดงผลข้อมูลมีความเป็นรูปแบบและเป็ นมาตรฐานเดียวกันยอมรับความพึงพอใจการใช้งานอยู่ในระดับดี 4) ในการเข้าใช้งานการเตือนเมื่อพบข้อผิดพลาดในกรณีที่ ผู้ใช้ไม่ป้อนข้อมูลตามที่กําหนดความเหมาะสมของระบบในการรักษาความปลอดภัยยอมรับความ พึงพอใจการใช้งานอยู่ในระดับดี

**ปัญหา :** ระบบได้กําหนดให้ผู้ใช้งานทั่วไป สามารถทําการโหวตให้คะแนนหนังสืออย่างน้อย 5เล่มขึ้นไปเพราะถ้าลงคะแนนให้กับหนังสือต่ำกว่า 5 เล่ม ผลของการแนะนําหนังสือจะไม่ตรงกับ 58 ความต้องการของผู้ใช้ ซึ่งในทางกลับกันถ้าโหวตหนังสือย่างน้อย 5 เล่มขึ้นไป ระบบจะแนะนํา หนังสือได้ตรงกับความต้องการของผู้ใช้

**ปัญหา :** ในการพัฒนาระบบระบบแนะนําหนังสือคอมพิวเตอร์แบบออนไลน์โดยใช้เทคนิคการ กรองแบบอิงเนื้อหา ซึ่งต้องใช้เวลาในการจัดการข้อมูลหนังสือทั้งหมด 448 เรื่อง โดยทําการจัด หมวดหมู่ให้กับหนังสือแต่ละเรื่อง ตลอดจนถึงการจัดการข้อมูลในส่วนของผู้แต่ง โดยทําการ จําแนกผู้แต่งออกเป็น 2 กลุ่ม คือผู้แต่งที่ได้รับความนิยม ผู้แต่งที่ไม่ได้รับความนิยม ให้กับหนังสือ ทั้งหมด ซึ่งการดําเนินงานดังกล่าวมานี้ทําให้เสียเวลาเป็นอย่างมาก เพื่อที่จะได้มาซึ่งข้อมูลที่ สมบูรณ์ในการนํามาใช้กับเทคนิคการกรองแบบอิงเนื้อหา

**ปัญหา :** เทคนิคการกรองแบบอิงเนื้อหาจะเหมาะสมกับข้อมูลหนังสือประมาณ 600 เรื่อง แต่ถ้ามี ข้อมูลหนังสือมากกว่า 1000 เรื่อง ขึ้นไปจะทําให้ประสิทธิภาพการทํางานของระบบช้าลง ซึ่งส่งผล ให้ผู้ใช้งานต้องรอผลการแนะนําเป็นเวลานาน กว่าจะได้ผลลัพธ์ออกมา