

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหา

กระทรวงศึกษาธิการมีนโยบายที่จะปรับปรุงระบบการรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษาเพื่อให้สอดคล้องกับนโยบายปฏิรูปการศึกษาของประเทศโดยมีหลักการสำคัญ 3 ประการ คือ 1. นักเรียนควรอยู่ในห้องเรียนจนจบมัธยมศึกษาปีที่ 6 (ม.6) 2) นักเรียนแต่ละคนมีเพียง 1 สิทธิในการตอบรับในสาขาวิชาที่เลือกเพื่อความเสมอภาค 3) สถาบันอุดมศึกษาในเครือข่ายที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) ทุก แห่งจะเข้าระบบเคลียร์ริงเฮาส์เพื่อบริหาร 1 สิทธิของนักเรียน การรับสมัครของนักเรียนการรับสมัครบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (Thai university Central Admission System : TCAS)

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เป็นมหาวิทยาลัยของรัฐที่อยู่ในเครือข่ายที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) ต้องดำเนินการตามนโยบายการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (TCAS) โดยมีสำนักส่งเสริมวิชาการและงานทะเบียนเป็นหน่วยงานหลักในการรับสมัครคัดเลือกบุคคลเข้าศึกษาต่อในระดับปริญญาตรีร่วมกับคณะและวิทยาลัยฯ รวมถึงกองประสานพันธ์และฝ่ายแนะแนวการศึกษาประจำคณะดำเนินการประชาสัมพันธ์ข้อมูลการสมัครเพื่อให้ได้นักศึกษาที่มีคุณสมบัติตรงตามที่สาขาวิชากำหนด

โครงการนี้มีความมุ่งหมายที่สำคัญโดยการนำเทคนิคการจำแนกข้อมูล (Data classification) มาวิเคราะห์พฤติกรรมการณ์ยืนยันสิทธิ์เคลียร์ริงเฮาส์ (Clearing house) เข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (TCAS) เพื่อเป็นการหารูปแบบที่เกิดขึ้นจากข้อมูลและนำรูปแบบที่ได้มาวางแผนการคัดเลือกบุคคลเข้าศึกษาในระบบ (TCAS) มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ซึ่งมหาวิทยาลัยมีจำนวนแผนการรับเข้าศึกษา 6,565 คน มีจำนวนผู้สมัคร 12,368 คน มีจำนวนผู้ผ่านการคัดเลือกที่มีสิทธิ์ยืนยันสิทธิ์เคลียร์ริงเฮาส์ 7,734 คน และจำนวนการยืนยันสิทธิ์ในการเข้าศึกษาต่อในมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา 4,350 คน ซึ่งคิดเป็นร้อยละ 56.25 (%) ต่อจำนวนผู้มีสิทธิ์ในการ ยืนยันสิทธิ์ Clearing house จากจำนวนการยืนยันสิทธิ์ ซึ่งมีจำนวนที่น้อยเมื่อเทียบกับแผนการรับเข้าศึกษาส่งผลกระทบกับแผนการรับนักศึกษาใหม่ หากย้อนมองถึงจำนวนผู้สมัครซึ่งเป็นผู้ที่สนใจจะเข้าศึกษาในมหาวิทยาลัยซึ่งมีจำนวนมาก (จาก ข้อมูล 12,368 คน) แต่ปัญหาคือมหาวิทยาลัยไม่สามารถรับเข้าศึกษาได้ทั้งหมด

เนื่องจากมีจำนวนมากเกินแผนการรับเข้าศึกษาดังนั้นมหาวิทยาลัยไม่สามารถคาดการณ์ได้ว่า จะมีผู้ยืนยันสิทธิ์มากน้อยเท่าใดจากปัญหาดังกล่าวผู้จัดทำได้มีแนวคิดในการนำเทคนิคเหมืองข้อมูล (Data mining) จากข้อมูลการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระดับปริญญาตรี (รอบที่ 2 การรับแบบโควตาที่มีการสอบข้อเขียน หรือข้อสอบปฏิบัติ) ประจำปีการศึกษา 2563 ข้อมูล (Data mining) การจำแนกข้อมูล (Data classification) เพื่อวิเคราะห์หารูปแบบที่เกิดขึ้น จากข้อมูลการยืนยันสิทธิ์เคลียร์ริงเฮาส์ (Clearing house) สามารถนำรูปแบบ (Model) ที่มาใช้วางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี นำไปสู่การวางแผนประชาสัมพันธ์ต่อไปในอนาคต เพื่อให้สามารถคาดการณ์หรือพยากรณ์ จำนวนผู้ยืนยันสิทธิ์เคลียร์ริงเฮาส์ก่อนล่วงหน้าได้ ซึ่งจะทำให้สามารถวางแผนจำนวนผู้ผ่านการคัดเลือกหรือจำนวนที่มีสิทธิ์ในการยืนยันสิทธิ์เคลียร์ริงเฮาส์ ซึ่งจะคาดคะเนจำนวนผู้ที่ยืนยันสิทธิ์เข้าศึกษาให้ใกล้เคียงกับแผนการรับจริงมากที่สุด

1.2 วัตถุประสงค์

เพื่อวิเคราะห์สารสนเทศที่เกิดขึ้นจากข้อมูลนักเรียนที่มีสิทธิ์ในการยืนยันสิทธิ์เข้าศึกษาต่อ โดยเทคนิคเหมืองข้อมูล (Data mining) มาใช้วางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS นำไปสู่การวางแผนประชาสัมพันธ์

1.3 ประโยชน์ที่ได้รับ

ได้ผลการวิเคราะห์สารสนเทศที่เกิดขึ้นจากข้อมูลนักเรียนที่มีสิทธิ์ในการยืนยันสิทธิ์เข้าศึกษาต่อ โดยเทคนิคเหมืองข้อมูล (Data mining) มาใช้วางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS นำไปสู่การวางแผนประชาสัมพันธ์

1.4 ขอบเขต

1.4.1 ขอบเขตผู้วิเคราะห์ข้อมูล

1.4.1.1 Business Understanding ผู้วิเคราะห์ข้อมูลทำความเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางด้าไมน์นิ่งและวางแผนในการดำเนินการโดยการวิเคราะห์ข้อมูลรูปแบบความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลของข้อมูลนักศึกษา และข้อมูลการรับสมัครนักศึกษา

1.4.1.2 Data Understanding ผู้วิเคราะห์ข้อมูลเริ่มทำการเก็บรวบรวมข้อมูลจากสำนักส่งเสริมวิชาการและงานทะเบียนและตรวจสอบข้อมูลที่ได้ดูความถูกต้องและความครบถ้วนของข้อมูล และพิจารณาข้อมูลทั้งหมดว่าส่วนใดของข้อมูลที่จำเป็นต่อการวิเคราะห์ข้อมูล

1.4.1.3 Data Preparation ผู้วิเคราะห์ข้อมูลทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมาและคัดเฉพาะส่วนที่จำเป็นให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้โดยการทำให้ข้อมูลให้ถูกต้องโดยใช้กระบวนการ Data Cleaning เพื่อให้ได้ข้อมูลอยู่ในรูปแบบและมาตรฐานเดียวกัน และกำจัดข้อมูลที่ไม่ได้มาตรฐานไม่ครบถ้วนและไม่เกี่ยวข้องออก

1.4.1.4 Modeling คณะผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification ด้วยการสร้างโมเดล Decision Tree และ Random Forest เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล (Data mining)

1.4.1.5 Evaluation คณะผู้จัดทำได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเหมืองข้อมูล (Data mining) จากการสร้างโมเดล Decision Tree และ Random Forest เพื่อสรุปผลข้อมูลเพื่อให้ตรงกับวัตถุประสงค์ที่ตั้งไว้

1.4.1.6 Deployment คณะผู้จัดทำนำผลข้อมูลที่ทำวิเคราะห์แล้วแสดงผลข้อมูลบนเว็บไซต์โดยใช้ชุดคำสั่ง HTML และ CSS3 และนำเสนอข้อมูลทั้งหมดแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้ Power BI ขอบเขตผู้ใช้งานทั่วไปบนเว็บไซต์

1.4.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บไซต์

1.4.2.1 สามารถดูแดชบอร์ดข้อมูลสารสนเทศของนักศึกษาย้อนหลังปีการศึกษา 2558 – 2562

1.4.2.2 สามารถดูแดชบอร์ดข้อมูลสารสนเทศของนักศึกษาย้อนหลังแต่ละปีของนักศึกษา

1.4.2.3 สามารถดูแดชบอร์ดข้อมูลสารสนเทศของนักศึกษาแบ่งตามข้อมูลต่างๆ

1.4.2.4 สามารถดูขั้นตอนการวิเคราะห์ข้อมูลบนเว็บไซต์ได้

1.4.2.5 สามารถสร้างโมเดล Decision Tree บนเว็บไซต์ได้

1.4.3 ขอบเขตการประเมินความเหมาะสมการแสดงผลข้อมูล

1.4.3.1 จัดทำแบบสอบถามเพื่อวิเคราะห์ความเหมาะสมในการแสดงผลข้อมูลที่ได้จากการวิเคราะห์โดยกลุ่มประชากรจากตัวแทนคณะเขตพื้นที่และหน่วยงานที่เกี่ยวข้องกับการประชาสัมพันธ์หรือแนะแนว จำนวน 30 คน

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 คอมพิวเตอร์แล็ปท็อปส่วนตัวเครื่องที่ 1

CPU – Intel Core i5-8300H

GPU – NVIDIA GeForce GTX 1050Ti

RAM – 8GB DDR4

Display – 15.6"

Storage – HDD 1TB 5400RPM

OS – Windows 10 Home (64 Bit)

1.5.1.2 คอมพิวเตอร์เดสก์ท็อปส่วนตัวเครื่องที่ 2

CPU – AMD Ryzen 5 3600

GPU – RX580

RAM – 16GB DDR4

Display – 24"

Storage – SSD 240GB.

OS – Windows 10 Home (64 Bit)

1.5.2 Software

1.5.2.1 โปรแกรม Knime ใช้ในการทำเทคนิคเหมืองข้อมูล

1.5.2.2 โปรแกรม Weka ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

1.5.2.3 โปรแกรม Tableau Public ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ

1.5.2.4 โปรแกรม MS Power BI ใช้ในการแสดงผลได้ทั้งรูปแบบ ตารางสรุปผลด้วย Visualization กราฟในรูปแบบต่างๆ

1.5.2.5 โปรแกรม MS Excel ใช้จัดรูปแบบของข้อมูล

1.5.2.6 โปรแกรม FileZilla โปรแกรมที่ใช้ติดต่อกับ FTP Server เพื่ออัปโหลด (Upload) ไฟล์ (File)

1.5.2.7 ภาษา HTML ใช้ในการเขียนเว็บเพจ โดยใช้ Tag ในการกำหนดการแสดงผล HTML

1.5.2.8 ภาษา CSS ใช้ในการตกแต่งเว็บไซต์

1.5.2.9 โปรแกรม Visual Studio Code ใช้ในการแก้ไขและปรับแต่งโค้ด

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

1.7 ระยะเวลาในการดำเนินงาน

แผนการดำเนินการ	2563					
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.
1. ศึกษาและกำหนดความต้องการ	↔					
2. การรวบรวมข้อมูลจากส่วนกลาง คณะและเขตพื้นที่ (Data Collection) และทำความเข้าใจ (Data Understanding)		↔				
3. จัดระเบียบข้อมูลเพื่อเตรียมสำหรับการวิเคราะห์ (Data Preparation)			↔			
4. วิเคราะห์ข้อมูล (Data Analysis)				↔	↔	
5. แสดงผลการวิเคราะห์ข้อมูลในรูปแบบต่าง ๆ				↔	↔	
6. นำผลการวิเคราะห์ข้อมูลมาวางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS					↔	↔
7. จัดทำคู่มือการใช้งาน					↔	↔
8. จัดทำเอกสารประกอบโครงการ					↔	↔

ตารางที่ 1.1 ระยะเวลาในการดำเนินงาน

1.8 นิยามศัพท์เฉพาะ

1.8.1 ระบบสนับสนุนการตัดสินใจ (Decision Support System) หมายถึง ระบบสารสนเทศที่ช่วยในการจัดเตรียมข้อมูลจัดรูปแบบ และนำข้อมูลที่มีความจำเป็นมาช่วยในการตัดสินใจ

1.8.2 Classification หมายถึง กระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดเป็นการสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ล่วงหน้า และสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้แบบจำลองที่ได้อาจอยู่ในรูปแบบการตัดสินใจแบบต้นไม้ตัดสินใจ (Decision Tree)

1.8.3 กลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) หมายถึง การนำเอากลุ่มข้อมูลตัวอย่าง (Training set) มาทำการวิเคราะห์โดยใช้เทคนิคของ Classification เพื่อทำการเรียนรู้และทำการสร้างแบบจำลองที่สามารถอธิบายถึงลักษณะความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ภายในฐานข้อมูล ซึ่งแบบจำลองนี้จะมีลักษณะที่กลุ่มของข้อมูลถูกแจกแจงออกเป็น Class ต่างๆ ด้วย Classification Rule และ Class แต่ละ Class จะมีลักษณะเฉพาะกลุ่มที่สามารถสรุปออกมาเป็นรูปแบบความสัมพันธ์ได้ (พยุชน พานิชย์กุล, 2548)

1.8.4 กลุ่มข้อมูลสำหรับการทดสอบ (Testing Data) หมายถึง การนำข้อมูลทดสอบ (Test Data) มาทดสอบแบบจำลองเพื่อดูความถูกต้องของ Classification Rule ที่ถูกสร้างมาจากขั้นตอนการเรียนรู้ (Training) เพื่อพิจารณาว่า Classification Rule ที่ถูกสร้างขึ้นมีความเหมาะสมที่จะนำไปใช้กับกลุ่มข้อมูลใหม่ๆ หรือไม่

1.8.5 ต้นไม้ตัดสินใจ (Decision Tree) หมายถึง ต้นไม้ที่ใช้ในการสนับสนุนการตัดสินใจซึ่งมีลักษณะเป็นโครงสร้างต้นไม้หัวกลับที่มีรากอยู่ด้านบนและใบอยู่ด้านล่างสุดโดยที่ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดนั้นจะแสดงถึงการตัดสินใจบนข้อมูลของคุณสมบัติต่าง ๆ กิ่งของต้นไม้แสดงถึงค่าหรือผลลัพธ์ที่ได้จากการทดสอบ และใบซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจจะแสดงถึงกลุ่มของข้อมูล (Class) หรือผลลัพธ์โหนดที่อยู่บนสุดจะเรียกว่าโหนดราก (Root Node) (นฤพนธ์ ว่องประชาณุกุล, 2548)

1.8.6 อัลกอริทึม (Algorithm) หมายถึง ขั้นตอนวิธีกระบวนการในการสอน (Training) ต้นไม้ตัดสินใจให้เรียนรู้ข้อมูลและสร้างแบบจำลอง

1.9 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น คณะผู้จัดทำได้เล็งเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลผู้สมัครและข้อมูลนักศึกษาเพื่อสนับสนุนการตัดสินใจในการวางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS ด้วยการนำเทคนิคเหมืองข้อมูล (Data mining) การจำแนกข้อมูล (Data classification) และรูปแบบของแผนภูมิต้นไม้ (Decision tree) และวิธีการจำแนกข้อมูลแบบ Random Forest เพื่อวิเคราะห์หารูปแบบที่เกิดขึ้นจากข้อมูลการยืนยันสิทธิ์เคลียร์ริงเฮาส์ (Clearing house) สามารถนำรูปแบบ (Model) ที่ได้มาวางแผนการคัดเลือกบุคคล เข้าศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา นำไปสู่การวางแผนประชาสัมพันธ์ต่อไปในอนาคต