การใช้เทคนิคเหมืองข้อมูลในการวิเคราะห์สารสนเทศเพื่อการพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบคัดเลือกบุคคลเข้าศึกษาต่อในระดับอุดมศึกษา

Application of data mining techniques in analysis of information to develop decision support systems in selecting programs under Thai University Central Admission System

อนันต์ ปินะเต[1]
Anan Pinate[1]

Received: 3 July 2020 ; Revised: 17 August 2020 ; Accepted: 1 September 2020

บทคัดย่อ

การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ หรือเรียกสั้นๆ ว่า TCAS เป็นนโยบายการปฏิรูปการศึกษาของกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรม (อว.) โดยมีที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) เป็นหน่วยงานกลางในการดำเนินงาน ซึ่งมหาวิทยาลัยทุกมหาวิทยาลัยเข้าร่วมการรับสมัครระบบใหม่นี้ ซึ่งมีรูปแบบการรับสมัครคัดเลือก 5 รอบ นโยบายนี้ได้ดำเนินการตั้งแต่ปีการศึกษา 2561 เป็นต้นมา มหาวิทยาลัยมหาสารคาม ได้เข้าร่วมการคัดเลือกระบบใหม่นี้ในทุกปีการศึกษา และทุกรอบการรับสมัคร ซึ่งทุกปีการศึกษามีผู้สมัครจำนวนมากที่ไม่ผ่านการคัดเลือกในสาขาวิชาเพื่อยืนยันสิทธิ์ (Clearing house) เข้าศึกษาในระบบ TCAS ได้ จากปัญหาพบว่าผู้สมัครไม่คำนึงถึงคะแนนการสอบวิชาความถนัดทั่วไป (GAT) และคะแนนการสอบความถนัดทางวิชาการและวิชาชีพ (PAT) ซึ่งเป็นองค์ประกอบหลักในการพิจารณาของสาขาวิชา โดยพฤติกรรมของผู้สมัครจะเลือกสาขาวิชาตามความรู้สึก ความชอบ และตามเพื่อน หรือผู้ปกครอง ปัญหาดังกล่าวผู้วิจัยได้นำเทคนิคเหมืองข้อมูล ด้วยวิธีต้นไม้ตัดสินใจ มาวิเคราะห์ข้อมูลและนำผลการวิจัยไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจเลือกสาขาวิชา เพื่อให้ผู้สมัครได้ทดสอบการเลือกสาขาวิชาเพื่อวิเคราะห์โอกาสผ่านการคัดเลือกเข้าศึกษามากน้อยเพียงใดก่อนเลือกสมัครสาขาวิชานั้นจริง จากการวิจัยพบว่าการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ C4.5 มีค่าความถูกต้องร้อยละ 91.64 สามารถสร้างกฎการตัดสินใจได้ทั้งหมด 358 กฎการตัดสินใจจากทั้งหมด 84 สาขาวิชา โดยสามารถนำกฎการตัดสินใจที่ได้มาพัฒนาเป็นระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS มหาวิทยาลัยมหาสารคาม

คำสำคัญ: ต้นไม้ตัดสินใจ การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (TCAS)

Abstract

Thai University Central Admission System, abbreviated to “TCAS,” is an educational reform policy under the Ministry of Higher Education, Science, Research and Innovation (MHESI), together with the Council of University Presidents of Thailand (CUPT) as the hub of operations. All universities have joined TCAS following five admissions application rounds. This policy has been implemented since the academic year of 2018. Mahasarakham University has participated in TCAS in every academic year and all admissions application rounds. Every year numerous candidates failed admissions into programs chosen for clearing house entering the TCAS system. There are problems that candidates overlooked scores on the General Aptitude Test (GAT) and Professional and Academic Aptitude Test (PAT) as a key component in a consideration of the program chosen. The candidate’s selection behavior toward a program is typically based on his/her feelings, preferences, and peers, and parent’s determination. In this study, the Data Mining technique was adopted, and Decision Tree was used to analyze the data. A result of the analysis was further used to develop a decision support system (DSS) for the program to be chosen by the applicants to test the program selection and to

[1] นักวิชาการคอมพิวเตอร์ชำนาญการ กองบริการการศึกษา มหาวิทยาลัยมหาสารคาม อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150
[1] Computer Techinical Officer Professional Level, Division of Academic Affair, Mahasarakham University, Kantharawichai District, MahaSarakham 44150 Thailand.

analyze the opportunity to achieve before the actual selection of the respective program. The results found that the Decision Tree at C4.5 represented an accuracy of 91.64 percent, and a total of 358 decision rules can be developed across 84 programs, and decision rules can be developed as a decision support system for program selection to the TCAS of Mahasarakham University.

Keywords: Decision tree, TCAS

บทนำ

กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรม (อว.) ได้มีนโยบายการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (Thai university Central Admission System: TCAS) (สมาคมที่ประชุมอธิการบดีแห่งประเทศไทย, 2561) ซึ่งได้ดำเนินการมาตั้งแต่ปีการศึกษา 2561 ซึ่งมีการคัดเลือกทั้งหมด 5 รอบด้วยกัน ได้แก่ รอบที่ 1 การรับด้วย Portfolio คือ การรับด้วยแฟ้มสะสมผลงานโดยไม่มีการสอบข้อเขียน กลุ่มผู้สมัครสำหรับนักเรียนทั่วไป นักเรียนที่มีความสามารถพิเศษ นักเรียนโควตา นักเรียนเครือข่าย ผู้สมัครสามารถสมัครได้ที่สถาบันอุดมศึกษา, รอบที่ 2 การรับแบบโควตาที่มีการสอบข้อเขียนหรือข้อสอบปฏิบัติ คือ การรับแบบโควตาที่มีการสอบข้อเขียนหรือข้อสอบปฏิบัติ สำหรับนักเรียนที่อยู่ในเขตพื้นที่หรือภาคโควตาโรงเรียนในเครือข่าย และโครงการความสามารถพิเศษต่าง ๆ ซึ่งสถาบันอุดมศึกษาประกาศเกณฑ์การสอบ ผู้สมัครสามารถสมัครได้ที่สถาบันอุดมศึกษาและเข้ารับการคัดเลือกตามเกณฑ์การสอบ, รอบที่ 3 การรับตรงร่วมกัน คือ การรับนักเรียนในโครงการกลุ่มสถาบันแพทยศาสตร์แห่งประเทศไทย (กสพท.) โครงการอื่นๆ และนักเรียนทั่วไปที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) เป็นหน่วยงานกลางในการรับสมัคร, รอบที่ 4 การรับกลางร่วมกัน Admission คือ การรับสำหรับนักเรียนทั่วไป ที่ประชุมอธิการบดีแห่งประเทศไทย เป็นหน่วยงานกลางในการรับสมัคร, รอบที่ 5 การรับตรงอิสระ คือ การรับโดยตรงด้วยวิธีการของสถาบันอุดมศึกษา และสถาบันอุดมศึกษาเป็นหน่วยงานรับสมัคร

มหาวิทยาลัยมหาสารคาม ได้เข้าร่วมการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (TCAS) ทั้ง 5 รอบที่ผ่านมาตั้งแต่ปีการศึกษา 2561-2563 (กองบริการการศึกษา, 2563: เว็บไซต์) ที่ผ่านมาพบว่ามีผู้จำนวนผู้สมัครจำนวนมากที่ไม่ผ่านการคัดเลือกในสาขาวิชาที่ผู้สมัครเลือกเนื่องจากมีคะแนนต่ำกว่าเกณฑ์การพิจารณาของสาขาวิชา จากรายงานสรุปผลการดำเนินงานการคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS (กองบริการการศึกษา, 2563) ประจำปีการศึกษา 2562 พบว่ามีผู้สมัครทั้งหมด 5 รอบมีผู้สมัครจำนวน 35,527 คน และมีจำนวนผู้ผ่านการคัดเลือกเพื่อยืนยันสิทธิ์เข้าศึกษา (Clearing House) 16,145 คน คิดเป็นร้อยละ 45.44 ของผู้ผ่านการคัดเลือก จากการวิเคราะห์ข้อมูลการสมัครของกองบริการการศึกษายังพบว่าพฤติกรรมการเลือกสาขาวิชาของผู้สมัครโดยส่วนใหญ่ยังขาดประสบการณ์ในการสมัครซึ่งผู้สมัครส่วนใหญ่จะเลือกสาขาวิชาตามความชอบและสภาพแวดล้อมอื่นๆ เช่น สมัครตามเพื่อน สมัครตามผู้ปกครองต้องการ (ชิดชนก ส่งศิริ, 2544) โดยไม่ได้คำนึงถึงความรู้ และทักษะของตนเอง เช่น ผลคะแนนการทดสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET), ผลคะแนนการทดสอบความถนัดทั่วไป (GAT), ความถนัดทางวิชาการและวิชาชีพ (PAT) (สถาบันทดสอบทางการศึกษาแห่งชาติ องค์การมหาชน, 2563: เว็บไซต์) และผลการทดสอบ 9 วิชาสามัญ ซึ่งคะแนนเหล่านี้คือองค์ประกอบหลักในการพิจารณาผลคัดเลือก ซึ่งส่งผลให้การสมัครคัดเลือกระบบ TCAS มหาวิทยาลัยมหาสารคาม ข้อมูลการรับสมัครประจำปีการศึกษา 2562 จากรายงานสรุปผลการดำเนินงานคัดเลือกบุคคลเข้าศึกษาในอุดมศึกษาระบบ TCAS มหาวิทยาลัยมหาสารคาม ประจำปีการศึกษา 2562 มีผู้ไม่ผ่านการคัดเลือกคิดเป็นร้อยละ 55.00 ของผู้สมัครทั้งหมด จากรายงานข้อมูลผู้ไม่ผ่านการคัดเลือกมีคะแนนไม่ถึงตามจำนวนสาขารับผ่านการคัดเลือก

ดังนั้นจากข้อมูลข้างต้น ทำให้ผู้วิจัยได้มีแนวคิดในการใช้เทคนิคเหมืองข้อมูล (Data Mining) ในการวิเคราะห์สารสนเทศเพื่อการพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาวิชา โดยระบบจะให้ผู้สมัครได้ทดสอบการเลือกสาขาวิชาก่อน เพื่อให้ทราบผลคะแนนของตน ว่ามีโอกาสมากหรือน้อยเพียงใด และวิเคราะห์โอกาสที่จะผ่านการคัดเลือกในสาขาวิชาที่ต้องการสมัคร ก่อนการเลือกสาขาวิชาจริง

วัตถุประสงค์

เพื่อวิเคราะห์สารสนเทศที่เกิดขึ้นจากข้อมูลการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS มหาวิทยาลัยมหาสารคาม โดยใช้เทคนิคเหมืองข้อมูล (Data Mining) และนำผลการทดลองมาพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS มหาวิทยาลัยมหาสารคาม

ทฤษฎีที่เกี่ยวข้อง

งานวิจัยนี้ผู้วิจัยได้นำทฤษฎีการทำเหมืองข้อมูลโดยวิธีการจำแนกข้อมูล (Data Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ในการวิเคราะห์ข้อมูล เพื่อให้ได้

สารสนเทศที่จะนำมาพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS รายละเอียดทฤษฎีเทคนิคเหมืองข้อมูล ดังต่อไปนี้

1. การทำเหมืองข้อมูล

เหมืองข้อมูล (Data mining) เป็นขั้นตอนทางสถิติเป็นวิธีการปัญญาประดิษฐ์ (Artificial intelligence) (ฉัตรเกล้า เจริญผล, 2556) ทำการวิเคราะห์และสกัดความรู้จากข้อมูลที่จัดเก็บไว้ในระบบฐานข้อมูล หรือรูปแบบข้อมูลอื่นๆ การทำเหมืองข้อมูลสามารถค้นพบความรู้ และสารสนเทศที่ซ่อนอยู่ในข้อมูล หลักการทำเหมืองข้อมูล คือ การวิเคราะห์หาความสัมพันธ์ของข้อมูล แนวโน้ม หรือหารูปแบบที่เกิดขึ้นจากข้อมูล และสามารถนำสารสนเทศที่ได้มาช่วยวางแผนการตัดสินใจและแก้ปัญหาด้านต่างๆ การทำเหมืองข้อมูลจะสามารถแก้ไขปัญหาได้บางปัญหาเท่านั้นขึ้นอยู่กับเทคนิควิธีการที่เลือกใช้เท่านั้น

การวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลมีกระบวนการตามมาตรฐานที่เรียกว่า Cross-Industry Standard Process for Data mining: CRISP-DM) (สมาคมที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.), 2561 ; สถาบันทดสอบทางการศึกษาแห่งชาติ องค์การมหาชน (สทศ.), 2563) เกิดขึ้นเมื่อปี ค.ศ. 1996 จากบริษัท Daimler Chysler บริษัท SPSS และบริษัท NCR เพื่อกำหนดมาตรฐานกลางในการทำเหมืองข้อมูลให้มีประสิทธิภาพ มีความน่าเชื่อถือในกระบวนการวิเคราะห์ข้อมูล กระบวนการ CRISP-DM รายละเอียดดัง Figure 1

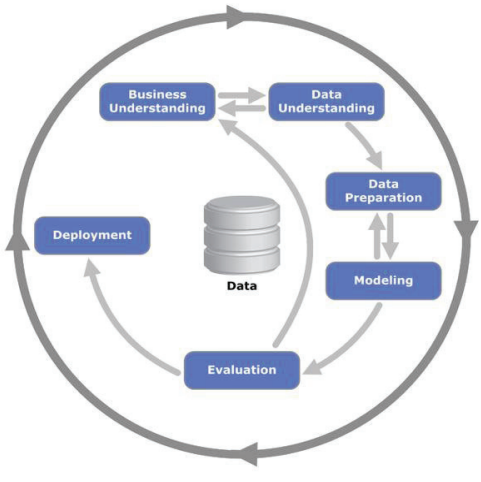

Figure 1 CRISP-DM

จาก Figure 1 แสดงกระบวนการการวิเคราะห์ข้อมูล CRISP-DM มีขั้นตอนการทำงาน 6 ขั้นตอน (Hand & Kamber, 2001) ได้แก่ 1) ขั้นตอนความเข้าใจธุรกิจ (Business Understanding) เป็นขั้นตอนการทำความเข้าใจ ปัญหาหรือโอกาสของธุรกิจ โดยทำการศึกษาปัญหาให้อยู่ในรูปแบบที่เหมาะสมในการวิเคราะห์เหมืองข้อมูล, 2) ขั้นตอนศึกษาความเข้าใจของข้อมูล (Data understanding) เป็นขั้นตอนของการรวบรวมข้อมูลที่เกี่ยวข้อง เพื่อใช้ในการวิเคราะห์เหมืองข้อมูล ข้อมูลที่นำมาวิเคราะห์ต้องมีความถูกต้อง น่าเชื่อถือ และมีความเหมาะสมในการวิเคราะห์เหมืองข้อมูล, 3) ขั้นตอนการเตรียมข้อมูล (Data preparation) เป็นขั้นตอนเตรียมข้อมูลตามรูปแบบที่ใช้ในการวิเคราะห์ ขั้นตอนการเตรียมข้อมูลเพื่อการวิเคราะห์ ได้แก่ การคัดเลือกข้อมูล การคัดกรองข้อมูล การแปลงข้อมูล ซึ่งขั้นตอนการเตรียมข้อมูลจะใช้เวลาพอสมควร, 4) ขั้นตอนการสร้างตัวแบบ (Modelling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยวิธีการต่างๆ ซึ่งการเลือกวิธีการต้องมีความหมาะสมในการวิเคราะห์ข้อมูล เพื่อให้ได้สารสนเทศที่มีประสิทธิภาพ ผู้วิเคราะห์ต้องมีการประเมินความถูกต้องของวิธีการต่างๆ ที่ใช้วิเคราะห์ เพื่อให้เกิดความน่าเชื่อถือ, 5) ขั้นตอนการประเมินผล (Evaluation) เป็นขั้นตอนในการประเมินผลที่ได้จากการวิเคราะห์เหมืองข้อมูลในรูปแบบต่างๆ ผลลัพธ์ที่ได้ตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่ เพื่อให้ได้ผลลัพธ์ที่ถูกต้องก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้, 6) ขั้นตอนการนำไปใช้ (Deployment) เป็นขั้นตอนการนำผลลัพธ์ที่ได้มาประยุกต์ใช้งาน การนำเสนอผลลัพธ์ที่ได้อาจนำมาแสดงเป็นรูปแบบรายงาน (Report) หรือพัฒนาเป็นระบบเพื่อความสะดวกต่อการใช้งานของผู้ใช้

2. การจำแนกข้อมูล

การจำแนกข้อมูล (Data Classification) (อนันต์ ปินะเต, 2559) เป็นวิธีการที่ใช้ในการค้นหาข้อมูลสารสนเทศความรู้ที่เกิดขึ้นจากข้อมูลในระบบฐานข้อมูลขนาดใหญ่ การจำแนกข้อมูลมีกระบวนการสร้างแบบจำลองการจำแนกข้อมูลเพื่อทำนายหมวดหมู่ของข้อมูล (Categories/Class) เช่น การวิเคราะห์ข้อมูลนักเรียนผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี มีโอกาสผ่านการคัดเลือกในสาขาวิชาที่สมัครหรือไม่ การวิเคราะห์ข้อมูลจะมีการสร้างตัวจำแนกข้อมูล จากชุดข้อมูลที่เป็นอินพุต ซึ่งแต่ละเรคคอร์ดประกอบไปด้วยเซตของแอทริบิวต์ที่บ่งบอกถึงคุณลักษณะของผู้สมัคร กระบวนการสร้างตัวจำแนกข้อมูลถูกเรียกว่า Learning หรือ Training ที่เกิดจากการนำเอาขั้นตอนวิธีการจำแนกข้อมูลมาดำเนินการกับข้อมูล ในชุดข้อมูลที่ทำการพิจารณาจะประกอบด้วยเซตของแอทริบิวต์ $x = (x_1, x_2,..., x_n)$ ที่บอกคุณลักษณะต่างๆ ของข้อมูลเรคคอร์ด x นอกจากนั้นเรคคอร์ด x ยังมีข้อมูลอีกหนึ่งแอทริบิวที่บ่งบอกถึงหมวดหมู่ของข้อมูล (Class lable attribute) โดยแอทริบิวต์หมวดหมู่ข้อมูลจะเป็นข้อมูลแบบไม่ต่อเนื่อง (Discrete-valued) โดยชุดข้อมูลที่เป็นพุตสำหรับการสร้างตัวจำแนกข้อมูลจะถูกเรียกว่าชุดข้อมูลสำหรับสอน (Training data) การจำแนกข้อมูลลักษณะนี้เรียกว่า การเรียนรู้แบบมีผู้สอน (Supervised learning)

3. การสร้างต้นไม้ตัดสินใจ

การสร้างต้นไม้ตัดสินใจ (Decision tree induction) (อนันต์ ปินะเต, 2559) เป็นกระบวนการสร้างต้นไม้ขึ้นเพื่อใช้ในการตัดสินใจจากข้อมูลที่มีหมวดหมู่ข้อมูล ต้นไม้ตัดสินใจจะประกอบไปด้วย โหนดต่างๆ ซึ่งถูกใช้ในการแสดงถึงเงื่อนไขหรือแอทริบิวต์หนึ่งๆ ของข้อมูล โดยที่แต่ละกิ่งก้านของโหนดหนึ่งๆ จะหมายถึงค่าที่เป็นไปได้จากการทดสอบกับแอทริบิวต์นั้นๆ และจะประกอบไปด้วยโหนดใบ (Leaf node) ซึ่งจะมีหมวดหมู่ข้อมูลจัดเก็บอยู่ โหนดการตัดสินใจนั้นจะมีการสร้างฟังก์ชันสำหรับการทดสอบ $f_m(x)$ การทดสอบจากการส่งข้อมูลเข้าจะทดสอบตามทางเลือกไปจนถึงโหนดใบและจะได้คำตอบจากการค้นหา รายละเอียดดัง Figure 2

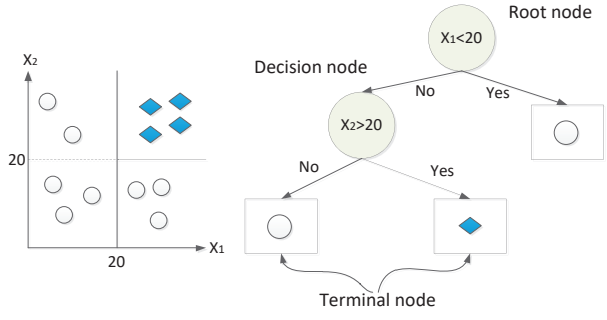

Figure 2 Tree map built from learning platform

อัลกอริทึม C4.5 (ชุติมา อุตมะมุณีย์, 2553: เว็บไซต์) เป็นอัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจ จะใช้ค่ามาตรฐานเกนในการตัดสินใจเลือกคุณสมบัติที่ใช้เป็นรากหรือโหนดในต้นไม้ โดยการคำนวณจากการเลือกคุณสมบัติที่มีค่าเกนสูงที่สุดมาเป็นโหนดรากการหาค่าเกนต้องมีการหาค่า Entropy, ค่า Information gain และค่า Split Information วิธีการหาค่ามีดังนี้

ค่า Entropy เป็นการหาค่าสารสนเทศของข้อมูล (Entropy measure) รายละเอียดดังสมการ 1

$$Entropy(S) = \sum_{i=1}^{c} -P_i \log_2 P_i \qquad (1)$$

โดย

S คือ แอทริบิวต์ที่นำมาวัดค่า

P_i คือ สัดส่วนของจำนวนสมาชิกในกลุ่ม i เท่ากับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

ค่า Information gain เป็นการหาค่าสารสนเทศก่อนนำไปใช้ในการหาค่ามาตรฐานอัตราส่วน เกน (Gain ratio) รายละเอียดดังสมการ 2

$$Gain(S, A) = Entropy(S) - \sum_{v=Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (2)$$

โดย

A คือ แอทริบิวต์ A

$|S_v|$ คือ สมาชิกของแอทริบิวต์ A

$|S|$ คือ จำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

ค่า Split Information เป็นการหาค่าสารสนเทศของการแบ่งข้อมูลใหม่ รายละเอียดดังสมการ 3

$$Split\ Information(S, A) = -\sum_{i=1}^{n} \frac{|S_i|}{S} \log_2 \frac{|S_i|}{|S|} \qquad (3)$$

โดย

S_i คือ จำนวนของจำนวนสมาชิกในกลุ่ม i

ค่า Gain ratio เพื่อใช้ในการลดทอนความเอนเอียงลงที่เกิดจากความโน้มเอียงจากการใช้ค่าเกน รายละเอียดดังสมการ 4

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{Split\ Information(S, A)} \qquad (4)$$

4. การวัดประสิทธิภาพ

การจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (อนันต์ ปินะเต, 2559) โดยใช้ค่าความถูกต้องของแบบจำลอง (Accuracy) ค่าความแม่นยำของแบบจำลอง (Precision) ค่าความระลึกของแบบจำลอง (Recall) และค่าความเหวี่ยงของแบบจำลอง (F-Measure) การวัดประสิทธิภาพของการจำแนกข้อมูลตามแนวคิดทางการค้นคืนสารสนเทศโดยอาศัยตาราง Confusion Matrix (สุพัฒน์กุล ภัคโชค, 2556) ในการคำนวณค่ารายละเอียดดัง Figure 3

		Predicted	
Actual	Class	Yes	No
	Yes	TP	FN
	No	FP	TN

Figure 3 Confusion Matrix

TP คือ จำนวนข้อมูลแบบจำลองการจำแนกกลุ่ม Yes และคำตอบเป็นกลุ่ม Yes

TN คือ จำนวนข้อมูลแบบจำลองการจำแนกกลุ่ม No และคำตอบเป็นกลุ่ม No

FP คือ จำนวนข้อมูลแบบจำลองการจำแนกกลุ่ม Yes และคำตอบเป็นกลุ่ม No

FN คือ จำนวนข้อมูลแบบจำลองการจำแนกกลุ่ม No และคำตอบ Yes

ค่าความถูกต้อง (Accuracy) คือ ค่าที่บอกถึงความถูกต้องของค่าที่วัดได้ใกล้เคียงความจริง รายละเอียดดังสมการ 5

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

ค่าความแม่นยำ (Precision) คือ ค่าสัดส่วนของข้อมูลที่ถูกต้อง และตรงตามความต้องการที่ถูกค้นคืนทั้งหมด หรือเป็นค่าที่แสดงถึงการค้นคืนสารสนเทศและมีความเป็นไปได้ตรงตามความต้องการมากน้อยเพียงใด รายละเอียดดังสมการ 6

$$\Pr ecision = \frac{TP}{TP + FP} \quad (6)$$

ค่าความระลึก (Recall) คือ ค่าสัดส่วนของข้อมูลที่ตรงตามความต้องการที่ถูกค้นคืนกับข้อมูลที่ต้องการทั้งหมด หรือค่าที่แสดงถึงความครอบคลุมในการสืบค้นข้อมูลสารสนเทศที่เกี่ยวข้องกับสิ่งที่ผู้ใช้ต้องการมากที่สุด รายละเอียดดังสมการ 7

$$\Re call = \frac{TP}{TP + FN} \quad (7)$$

ค่าความเหวี่ยง (F-Measure) คือ ค่าเฉลี่ยค่าความแม่นยำในการตรวจพบ และค่าความระลึกเป็นการวัดค่าความแม่นยำโดยรวม รายละเอียดดังสมการ 8

$$F - Measure = \frac{2 * \Pr ecision * \Re call}{\Pr ecision + \Re call} \quad (4)$$

วิธีการดำเนินงานวิจัย

งานวิจัยนี้ได้นำเสนอการใช้ข้อมูลการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS มหาวิทยาลัยมหาสารคาม ปีการศึกษา 2561-2563 โดยการใช้เทคนิคเหมืองข้อมูล (Data mining) ด้วยวิธีการจำแนกข้อมูลแบบต้นไม้ตัดสินใจ (Decision tree) เพื่อค้นหาสารสนเทศที่ซ่อนอยู่ในข้อมูล และนำรูปแบบที่ได้มาพัฒนาเป็นระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS เพื่อให้นักเรียนผู้ที่สนใจเข้าศึกษา ร่วมถึงครูแนะแนวหรือผู้ปกครองสามารถทดลองใช้ระบบสนับสนุนการเลือกสาขาวิชาเพื่อให้ทราบโอกาสในการผ่านเข้าศึกษาในสาขาวิชาที่ต้องการสมัคร ก่อนการสมัครในสาขาวิชานั้นจริง ขั้นตอนการดำเนินงานและกรอบแนวคิดการวิจัยมีรายละเอียดดัง Figure 4

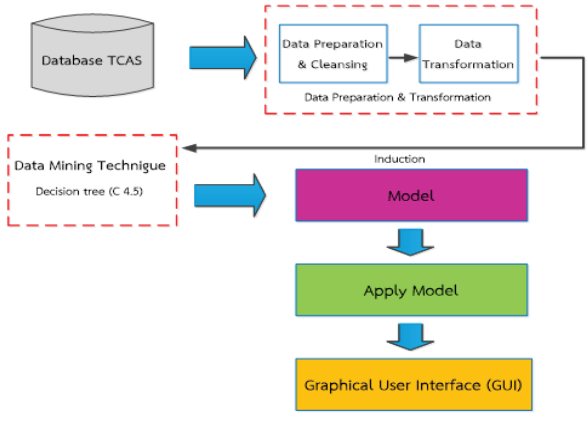

Figure 4 Conceptual framework

การวิจัยนี้ได้เลือกข้อมูลที่เกี่ยวข้องกับการพิจารณาผลการผ่านการคัดเลือกเข้าศึกษาในระบบ TCAS มหาวิทยาลัยมหาสารคาม ซึ่งได้แก่ ข้อมูลผลคะแนนการทดสอบความรู้ทั่วไป (General Aptitude Test: GAT) และผลคะแนนการทดสอบความรู้ทางวิชาการและวิชาชีพ (Professional and Academic Aptitude Test: PAT) รายละเอียดผลคะแนนวิชาของการทดสอบ GAT/PAT ที่ใช้เป็นข้อมูลทดลองมีรายละเอียด ดังนี้ 1) GAT (ตอนที่ 1) คือคะแนนความสามารถในการอ่าน การเขียน การคิดวิเคราะห์ และการแก้โจทย์ปัญหา, 2) GAT (ตอนที่ 2) คือคะแนนความสามารถในการสื่อสารด้านภาษาอังกฤษ, 3) GAT (รวม) คือคะแนนรวมทั้ง 2 ตอน, 4) PAT1 คือคะแนนความถนัดทางคณิตศาสตร์, 5) PAT2 คือคะแนนความถนัดทางวิทยาศาสตร์, 6) PAT3 คือคะแนนความถนัดทางวิศวกรรมศาสตร์, 7) PAT4 คือคะแนนความถนัดทางสถาปัตยกรรมศาสตร์, 8) PAT5 คือคะแนนความถนัดทางวิชาชีพครู, 9) PAT6 คือคะแนนความถนัดทางศิลปกรรมศาสตร์, 10) PAT7 คือคะแนนความถนัดทางภาษาต่างประเทศ และคะแนนรวม (Sum Score) คือคะแนนรวมตามองค์ประกอบของสาขาวิชา รายละเอียดตัวอย่างข้อมูล ดัง Table 1

Table 1 Data sample

GAT1	GAT2	GAT	PAT1	PAT2	PAT3	PAT4	PAT5	PAT6	PAT7	Sum Score
66.58	62.71	68.17	53.69	69.35	57.39	0.00	71.73	66.58	62.71	68.17
59.78	62.07	53.14	63.12	65.20	0.00	0.00	68.52	59.78	62.07	53.14
48.22	50.35	45.18	0.00	0.00	0.00	0.00	0.00	48.22	50.35	45.18
45.16	43.31	19.60	43.58	55.50	38.36	0.00	0.00	45.16	43.31	49.60
50.23	52.69	46.06	35.49	47.18	0.00	0.00	49.26	50.23	52.69	46.06
49.73	55.18	40.76	55.71	49.96	44.34	0.00	0.00	49.73	55.18	40.76
59.09	58.94	56.67	65.14	54.11	0.00	0.00	66.92	59.09	58.94	56.67
...	...	...	...	...	...	...	...	...	...	...

จาก Table 1 แสดงตัวอย่างข้อมูลที่ใช้ในการวิเคราะห์ข้อมูล จากข้อมูลผู้วิจัยได้ทำการเตรียมข้อมูลโดยการรวมข้อมูลที่ใช้วิเคราะห์และทำการเตรียมข้อมูลเพื่อให้พร้อมในการวิเคราะห์ตามอัลกอริทึมของการทำเหมืองข้อมูล วิธีต้นไม้ตัดสินใจ (Decision tree) ซึ่งข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลทศนิยมแบบต่อเนื่องและมีการกระจายของข้อมูล ซึ่งผู้วิจัยต้องทำการลดการกระจายของข้อมูลโดยการจัดทำกลุ่มข้อมูล (Binning data) โดยการแบ่งช่วงผลคะแนนเพื่อความเหมาะสมของการวิเคราะห์ข้อมูล งานวิจัยนี้ใช้กระบวนการแปลงคุณลักษณะของข้อมูลแบบต่อเนื่องให้อยู่ในรูปบบของคุณลักษณะข้อมูลแบบไม่ต่อเนื่องเพื่อช่วยในการลดขนาดและความซับซ้อนของข้อมูล โดยใช้หลักการแบ่งช่วงข้อมูล (Discretization) โดยการแบ่งช่วงข้อมูลตามกลุ่มดังนี้ กลุ่มที่ 1 กลุ่มคะแนน GAT (ตอนที่1), GAT (ตอนที่ 2), GAT (รวม), PAT1, PAT2, PAT3, PAT4, PAT5, PAT6 และ PAT7 จากคะแนนแต่ละวิชาทำการแบ่งระดับคะแนนออกเป็น 4 ระดับได้แก่ ระดับต่ำ คะแนนระหว่าง 0.00-39.08 แทนค่าเป็น (L), ระดับกลาง คะแนนระหว่าง 39.09-51.48 แทนค่าเป็น (M), ระดับสูง คะแนนระหว่าง 51.49-63.88 แทนค่าเป็น (H), ระดับสูงมาก คะแนนระหว่าง 63.89-100 แทนค่าเป็น (VH) และกลุ่มที่ 2 คะแนนรวม (Sum Score) แบ่งระดับคะแนนออกเป็น 4 ระดับได้แก่ ระดับโอกาสผ่านเข้าศึกษา น้อย คะแนนรวมระหว่าง 0.00-50.58 แทนค่าเป็น (Low), ระดับโอกาสผ่านเข้าศึกษา ปานกลาง คะแนนรวมระหว่าง 50.59-65.10 แทนค่าเป็น (Moderate), ระดับโอกาสผ่านเข้าศึกษา มาก คะแนนรวมระหว่าง 65.11-79.62 แทนค่าเป็น (High), ระดับโอกาสผ่านเข้าศึกษา มากสุด คะแนนรวมระหว่าง 79.63-100 แทนค่าเป็น (Most) จากการเตรียมข้อมูล การลดการกระจายของข้อมูลโดยจัดกลุ่มข้อมูล และกำหนดระดับของคะแนนสามารถแสดงรายละเอียดดัง Table 2

Table 2 Variable sample

GAT1	GAT2	GAT	PAT1	PAT2	PAT3	PAT4	PAT5	PAT6	PAT7	Sum Score
VH	H	VH	H	VH	H	L	VH	VH	H	Most
H	H	H	H	VH	L	L	VH	H	H	High
M	M	M	L	L	L	L	L	M	H	Low
M	M	M	M	H	L	L	L	M	M	Low
M	H	M	L	M	L	L	M	M	H	Low
M	H	M	H	M	M	L	L	M	H	Moderate
H	H	H	VH	H	L	L	VH	H	H	High
VH	VH	VH	H	H	M	L	L	L	L	Moderate
...	...	...	...	...	...	...	...	...	...	...

จาก Table 2 ตัวอย่างข้อมูลที่ใช้ในการวิเคราะห์ข้อมูลโดยการจำแนกข้อมูลแบบวิธีต้นไม้ตัดสินใจ งานวิจัยนี้มีการวัดประสิทธิภาพผลการวิเคราะห์ซึ่งวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) การแบ่งข้อมูลเป็น 2 กลุ่มคือ ข้อมูลทดสอบ (Data testing) และข้อมูลเรียนรู้ (Data training) โดยการแบ่งข้อมูลแยกตามสาขาวิชาเพื่อทำการวิเคราะห์ข้อมูล ซึ่งใช้หลักการแบ่งข้อมูลแบบ Cross validation เป็นการแบ่งข้อมูลออกเป็นการตรวจสอบแบบไขว้ โดยเลือกใช้แบบ 10 ส่วน (10 fold cross validation)

ขั้นตอนทำการสร้างโหนดสำหรับข้อมูลทั้งหมดในชุดข้อมูลแอทริบิวต์ จากนั้นทำการค้นหาแอทริบิวต์ที่ใช้ในการแยกข้อมูลออกเป็นชุดย่อยๆ ด้วยการคำนวณหาค่าเกนของแต่ละแอทริบิวต์ (GAT1, GAT2, GAT, PAT1, PAT2, PAT3, PAT4, PAT5, PAT6 และ PAT7) ดังสมการ 1

ขั้นตอนการคำนวณค่าเกนความรู้ของแต่ละแอทริบิวต์ โดยทำการพิจารณาแอทริบิวต์ GAT1 ที่มีค่า 4 ค่า คือ L, M, H และ VH ตามลำดับ หลังจากคำนวณค่าเกนของ GAT1 แล้วจะทำการคำนวณค่าเกนของแอทริบิวต์อื่นทั้งหมด

โปรแกรมที่ใช้ในงานวิจัยนี้ คือ โปรแกรม Weka (Waikato Environment for Knowledge Analysis) (Zdravko Markov, 2020: เว็บไซต์) ซึ่งเป็นโปรแกรมสำเร็จรูปประเภทฟรีโปรแกรม การวิเคราะห์ข้อมูลโดยการเลือกวิธีการจำแนก (Classify) เลือกวิธีต้นไม้ตัดสินใจ อัลกอริทึม C4.5 (J48) โปรแกรม Weka จะทำการวิเคราะห์ข้อมูลและจะแสดงผลการวิเคราะห์รายละเอียดดัง Figure 5

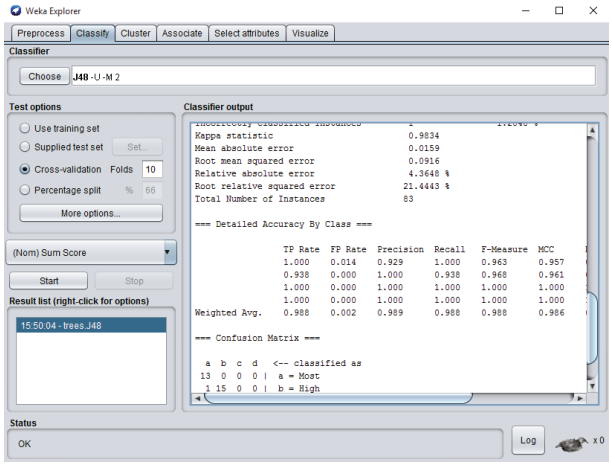

Figure 5 Analysis result

ขั้นตอนการสร้างรูปแบบการจำแนกข้อมูล (Classification model) และการนำแบบจำลองมาใช้มีรายละเอียดดัง Figure 6

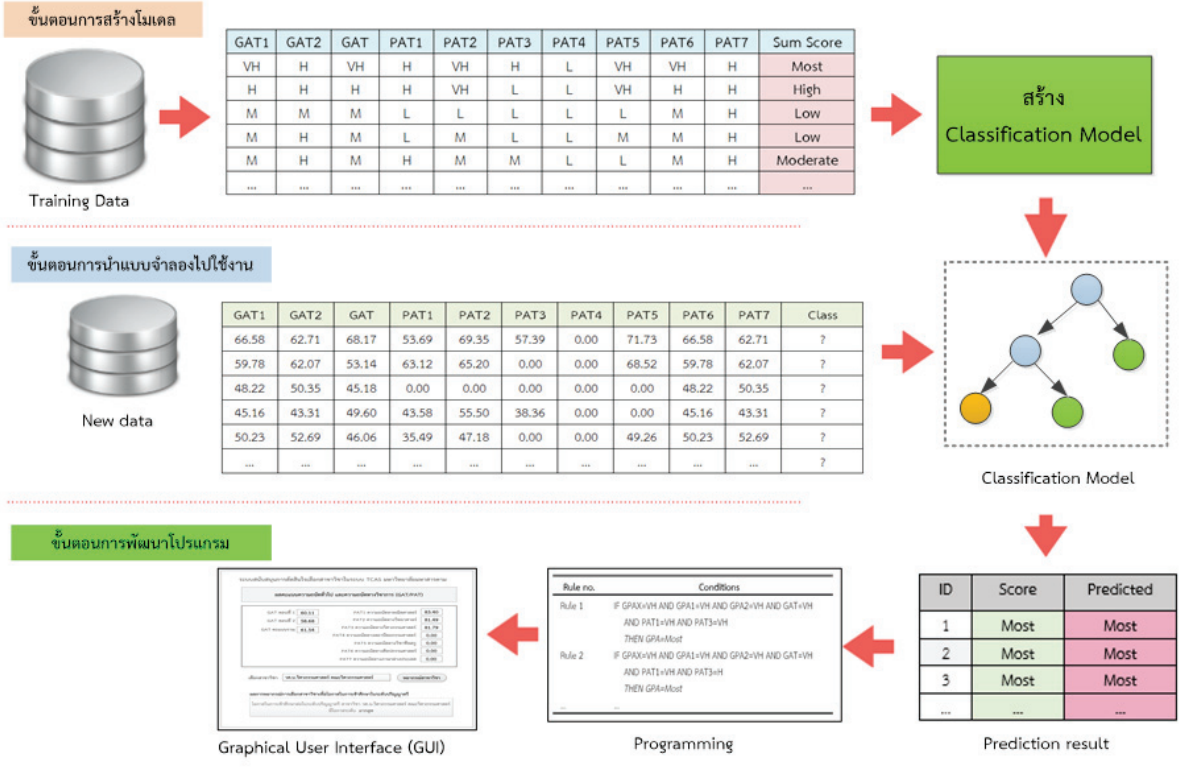

Figure 6 The Creation Classification model

ผลการศึกษาวิจัย

จากการวิเคราะห์ข้อมูลซึ่งผู้วิจัยได้ทำการวิเคราะห์ข้อมูลด้วยโปรแกรม Weka โดยทำการวิเคราะห์ในแต่ละสาขาวิชาด้วยวิธีต้นไม้ตัดสินใจ (Decision tree) แบบ C4.5 ผลการวิเคราะห์ข้อมูลในแต่ละสาขาวิชา ได้ผลการวิเคราะห์จากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความเหวี่ยง (F-Measure) รายละเอียดผลการวิเคราะห์ข้อมูลในแต่ละสาขาวิชาดัง Table 3

Table 3 Result

No.	Bachelor Program	Decision tree (C4.5)				
		Accuracy	Precision	Recall	F-Measure	Class
1	B.ATM. Applied Thai Traditional	87.31	0.880	0.873	0.871	High
2	B.Sc. Emergency Medicine Operation	88.10	0.885	0.881	0.877	High
3	Pharm.D. Pharmaceutical Care	74.78	0.740	0.748	0.740	High
4	B.N.S. Nursing Science	87.26	0.806	0.873	0.824	Moderate
5	B.P.H. Public Health	87.96	0.881	0.880	0.875	High
6	B.Sc. Nutritional Science Dietetics and Food Safety	90.04	0.898	0.900	0.896	Moderate
7	B.Sc. Environmental Health	90.45	0.907	0.905	0.897	Moderate
8	B.Sc. Occupational Health and Safety	91.49	0.915	0.915	0.908	Moderate
9	B.Sc. Chemistry	87.72	0.883	0.877	0.877	High
10	B.Sc. Biology	88.32	0.877	0.883	0.879	High
11	B.Sc. Physics	90.40	0.898	0.904	0.900	High
12	B.Sc. Applied Physics	82.46	0.824	0.825	0.820	High
13	B.Sc. Mathematics	86.51	0.872	0.865	0.863	High
14	B.Sc. Statistics	88.27	0.877	0.883	0.878	High
15	B.Sc. Microbiology	85.83	0.855	0.858	0.856	Moderate
16	B.Sc. Food technology and Nutrition	87.14	0.870	0.871	0.870	Moderate
17	B.Sc. Biotechnology	92.95	0.917	0.930	0.923	Moderate
18	B.Sc. Agriculture (Horticulture)	89.98	0.893	0.900	0.895	Moderate
19	B.Sc. Agriculture (Agronomy)	88.35	0.877	0.884	0.877	Moderate
20	B.Sc. Food Product Development	88.50	0.881	0.885	0.882	Moderate
21	B.Sc. Animal Science	90.52	0.904	0.905	0.904	High
22	B.Sc. Fisheries	90.73	0.909	0.907	0.901	Moderate
23	D.V.M. Veterinary Medicine	80.75	0.811	0.808	0.806	High
24	B.Sc. Environmental Technology	90.53	0.904	0.905	0.898	Moderate
25	B.Sc. Environmental and Natural Resources Management	92.19	0.915	0.922	0.918	Moderate
26	B.Sc. Environmental Education	91.05	0.922	0.911	0.911	Low
27	B.A. Information Science	100.0	1.000	1.000	1.000	Most
28	B.Sc. Information Technology	82.27	0.815	0.823	0.892	Moderate
29	B.Sc. Computer Science	89.79	0.893	0.898	0.895	Moderate
30	B.Sc. Creative Media	92.27	0.930	0.923	0.923	Moderate
31	B.Sc. Geo-Informatics	96.42	0.971	0.964	0.966	High

Table 3 Result (cont.)

No.	Bachelor Program	Decision tree (C4.5)				
		Accuracy	Precision	Recall	F-Measure	Class
32	B.Com. Arts Communication Arts	99.66	0.997	0.997	0.997	Most
33	B.Eng. Engineering	89.59	0.893	0.896	0.894	Moderate
34	B.Arch. Architecture	85.88	0.866	0.859	0.850	High
35	B.Arch. Urban Architecture	89.37	0.897	0.894	0.889	Moderate
36	B.Arch. Interior Architecture	83.74	0.841	0.837	0.827	Moderate
37	B.F.A. Creative Arts	82.40	0.803	0.824	0.810	High
38	B.L.A. Landscape Architecture	77.98	0.749	0.780	0.758	Moderate
39	B.Sc. Construction Management	99.01	0.981	0.990	0.985	Most
40	B.Acc. Accountancy	100.0	1.000	1.000	1.000	Most
41	B.B.A. Marketing	100.0	1.000	1.000	1.000	Most
42	B.B.A. Management	100.0	1.000	1.000	1.000	Most
43	B.B.A. Business Computer	100.0	1.000	1.000	1.000	Most
44	B.B.A. International Business (International Program)	100.0	1.000	1.000	1.000	Most
45	B.B.A. Financial Management	100.0	1.000	1.000	1.000	Most
46	B.B.A. Business Information Technology	100.0	1.000	1.000	1.000	Most
47	B.B.A. Human Resources Management	100.0	1.000	1.000	1.000	Most
48	B.Econ. Business Economics	100.0	1.000	1.000	1.000	Most
49	B.A. Tourism Management	100.0	1.000	1.000	1.000	Most
50	B.A. Hotel Management	100.0	1.000	1.000	1.000	Most
51	B.A. Tourism Management (English Program)	100.0	1.000	1.000	1.000	Most
52	B.A. Thai Language	100.0	1.000	1.000	1.000	Most
53	B.A. Language for Media Creation	99.68	0.997	0.997	0.997	Most
54	B.A. English	76.08	0.763	0.761	0.752	High
55	B.A. English for International Communication (International Program)	76.58	0.760	0.766	0.758	High
56	B.A. Business English	72.34	0.751	0.723	0.711	High
57	B.A. Chinese	79.40	0.839	0.794	0.790	Moderate
58	B.A. Japanese	100.0	1.000	1.000	1.000	Most
59	B.A. French	87.77	0.894	0.878	0.880	Most
60	B.A. History	100.0	1.000	1.000	1.000	Most
61	B.A. Community and Social Development	100.0	1.000	1.000	1.000	Most
62	B.A. Development Geography for Resources Management	99.39	0.994	0.994	0.994	High
63	B.A. Korean	96.71	0.970	0.967	0.968	Most
64	B.A. ASEAN Languages and Cultures (Khmer Language)	97.82	0.980	0.978	0.979	High
65	B.A. ASEAN Languages and Cultures(Lao Language)	97.05	0.971	0.971	0.971	High
66	B.A. ASEAN Languages and Cultures (Vietnamese Language)	96.20	0.963	0.962	0.962	High

Table 3 Result (cont.)

No.	Bachelor Program	Decision tree (C4.5)				
		Accuracy	Precision	Recall	F-Measure	Class
67	B.Pol.Sci. Political and Government	100.0	1.000	1.000	1.000	Most
68	B.Pol.Sci. Public Administration	100.0	1.000	1.000	1.000	Most
69	B.Pol.Sci. international Relations	100.0	1.000	1.000	1.000	Most
70	B.Ed. General Science	89.29	0.894	0.893	0.893	High
71	B.Ed. Mathematics	86.39	0.860	0.864	0.861	High
72	B.Ed. Social Studies	83.65	0.841	0.837	0.822	Moderate
73	B.Ed. English	82.04	0.814	0.820	0.817	High
74	B.Ed. Thai Language	86.33	0.858	0.863	0.854	Moderate
75	B.Ed. Early Childhood Education	79.92	0.811	0.799	0.788	Moderate
76	B.Ed. Educational Technology and Computer Education	86.93	0.865	0.869	0.860	Moderate
77	B.Sc. Psychology	93.83	0.945	0.938	0.940	Most
78	B.Sc. Sport Science	100.0	1.000	1.000	1.000	Most
79	B.F.A. Visual Arts	91.26	0.912	0.913	0.911	High
80	B.F.A. Performing Arts	96.91	0.969	0.969	0.969	Most
81	B.F.A. Product Design and Development	90.24	0.908	0.902	0.902	High
82	B.M. Music	94.31	0.941	0.943	0.941	Moderate
83	B.A. Cultural Management	100.0	1.000	1.000	1.000	High
84	L.L.B Laws	100.0	1.000	1.000	1.000	Most

ผลการวิเคราะห์ข้อมูลในแต่ละสาขาวิชา จำนวน 84 สาขาวิชา ด้วยวิธีต้นไม้ตัดสินใจ (Decision tree) แบบ C4.5 สามารถสร้างต้นไม้ และต้นไม้ตัดสินใจ (Decision rule) ในต้นไม้แต่ละสาขาวิชา ตัวอย่างต้นไม้ตัดสินใจ สาขาวิชา วท.บ. วิทยาการคอมพิวเตอร์ (B.Sc. Computer Science) มีลักษณะต้นไม้ตัดสินใจ ดัง Figure 7

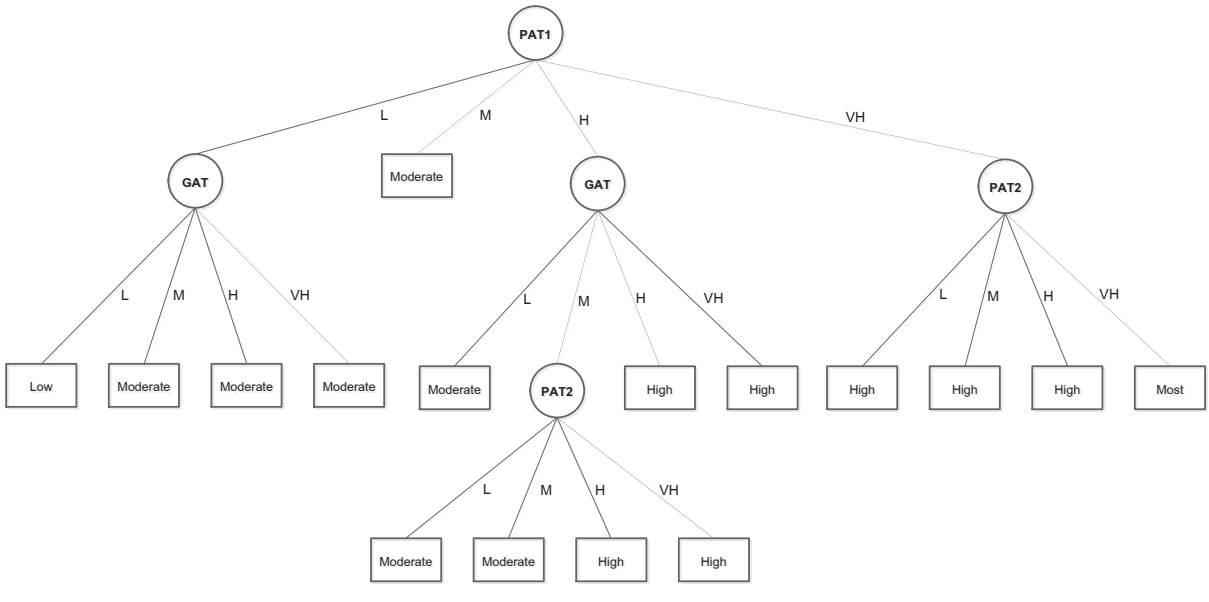

Figure 7 Example Tree

จาก Figure 7 ต้นไม้ที่ได้จากการวิเคราะห์จำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจของข้อมูลสาขาวิชา วท.บ. วิทยาการคอมพิวเตอร์ จากต้นไม้ตัดสินใจสร้างเป็นกฎการตัดสินใจ (Decision rule) ได้จำนวน 9 กฎ รายละเอียดกฎการตัดสินใจ ดังนี้

Rule1 **IF** PAT1=VH **AND** PAT2=VH

THEN Sum Score=Most

Rule2 **IF** PAT1=VH **AND** PAT2=(H **OR** M **OR** L)

THEN Sum Score=High

Rule3 **IF** PAT1=H **AND** GAT=(VH **OR** M)

THEN Sum Score=High

Rule4 **IF** PAT1=H **AND** GAT=M **AND** PAT2=(VH OR H)

THEN Sum Score=High

Rule5 **IF** PAT1=H **AND** GAT=M **AND** PAT2=(M OR L)

THEN Sum Score=Moderate

Rule6 **IF** PAT1=H **AND** GAT=L

THEN Sum Score=Moderate

Rule7 **IF** PAT1=M

THEN Sum Score=Moderate

Rule8 **IF** PAT1=L **AND** GAT=(VH OR H OR M)

THEN Sum Score=Moderate

Rule9 **IF** PAT1=L **AND** GAT=L

THEN Sum Score=Low

กฎข้อที่ 1 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับสูงมาก คะแนนระหว่าง 63.89-100 (VH) และคะแนนความถนัดทางวิทยาศาสตร์ (PAT 2) ระดับสูงมาก คะแนนระหว่าง 63.89-100 (VH) ผู้สมัครมีโอกาสผ่านการคัดเลือกระดับมากสุด (Most)

กฎข้อที่ 2 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับสูงมาก คะแนนระหว่าง 63.89-100 (VH) และคะแนนความถนัดทางวิทยาศาสตร์ (PAT 2) ระดับสูง คะแนนระหว่าง 51.49-63.88 (H) หรือระดับกลาง คะแนนระหว่าง 39.09-51.48 (M) หรือระดับต่ำ คะแนนระหว่าง 0.00-39.08 (L) ผู้สมัครมีโอกาสผ่านการคัดเลือกระดับมาก (High)

กฎข้อที่ 3 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับสูง คะแนนระหว่าง 51.49-63.88 (H) และคะแนนความถนัดทั่วไป (GAT) ระดับสูงมาก คะแนนระหว่าง 63.89-100 (VH) หรือระดับสูง คะแนนระหว่าง 51.49-63.88 (H) สมัครมีโอกาสผ่านการคัดเลือกระดับมาก (High)

กฎข้อที่ 4 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับสูง คะแนนระหว่าง 51.49-63.88 (H) และคะแนนความถนัดทั่วไป (GAT) ระดับกลาง คะแนนระหว่าง 39.09-51.48 (M) คะแนนความถนัดทางวิทยาศาสตร์ (PAT 2) ระดับสูงมาก คะแนนระหว่าง 63.89-100 (VH) หรือระดับสูง คะแนนระหว่าง 51.49-63.88 (H) สมัครมีโอกาสผ่านการคัดเลือกระดับมาก (High)

กฎข้อที่ 5 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับสูง คะแนนระหว่าง 51.49-63.88 (H) และคะแนนความถนัดทั่วไป (GAT) ระดับกลาง คะแนนระหว่าง 39.09-51.48 (M) และคะแนนความถนัดทางวิทยาศาสตร์ (PAT 2) ระดับกลาง คะแนนระหว่าง 39.09-51.48 (M) หรือระดับต่ำ คะแนนระหว่าง 0.00-39.08 (L) สมัครมีโอกาสผ่านการคัดเลือกระดับปานกลาง (Moderate)

กฎข้อที่ 6 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับสูง คะแนนระหว่าง 51.49-63.88 (H) และคะแนนความถนัดทั่วไป (GAT) ระดับต่ำ คะแนนระหว่าง 0.00-39.08 (L) สมัครมีโอกาสผ่านการคัดเลือกระดับปานกลาง (Moderate)

กฎข้อที่ 7 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับกลาง คะแนนระหว่าง 39.09-51.48 (M) สมัครมีโอกาสผ่านการคัดเลือกระดับปานกลาง (Moderate)

กฎข้อที่ 8 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับต่ำ คะแนนระหว่าง 0.00-39.08 (L) และคะแนนความถนัดทั่วไป (GAT) ระดับสูงมาก คะแนนระหว่าง 63.89-100 (VH) หรือระดับสูง คะแนนระหว่าง 51.49-63.88 (H) หรือระดับกลาง คะแนนระหว่าง 39.09-51.48 (M) สมัครมีโอกาสผ่านการคัดเลือกระดับปานกลาง (Moderate)

กฎข้อที่ 9 หากผู้สมัครเลือกสาขาวิชาวิทยาการคอมพิวเตอร์ มีคะแนนความถนัดทางคณิตศาสตร์ (PAT 1) ระดับต่ำ คะแนนระหว่าง 0.00-39.08 (L) สมัครมีโอกาสผ่านการคัดเลือกระดับน้อย (Low)

กฎการตัดสินใจที่ได้ในแต่ละสาขาวิชาผู้วิจัยสามารถนำมาเขียนโปรแกรมเพื่อพัฒนาเป็นระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS มหาวิทยาลัยมหาสารคาม รายละเอียดระบบสนับสนุนการตัดสินใจดัง Figure 8

Figure 8 Forecast TCAS

สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS โดยใช้เทคนิคเหมืองข้อมูล (Data mining) และผลการวิจัยมาพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS มหาวิทยาลัยมหาสารคาม

จากการวิจัย การนำเทคนิคเหมืองข้อมูล โดยการจำแนกข้อมูล (Data Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision tree) แบบ C4.5 ซึ่งข้อมูลที่ใช้วิเคราะห์ คือข้อมูลการรับสมัครคัดเลือกบุคคลเข้าศึกษาระบบ TCAS ปีการศึกษา 2561-2563 การประเมินผลการวิจัยด้วยค่าความถูกต้อง (Accuracy) หลักการแบ่งข้อมูลในการวิเคราะห์แบ่งข้อมูลออกเป็น 2 กลุ่ม คือข้อมูลทดสอบ (Data testing) และข้อมูลเรียนรู้ (Data training) ใช้หลักการการตรวจสอบข้อมูลแบบไขว้ Cross validation โดยเลือกใช้แบบ 10 ส่วน (10 fold cross validation)

จากการวิจัยพบว่าการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ มีค่าความถูกต้องร้อยละ 91.64 สามารถสร้างกฎการตัดสินใจได้ทั้งหมด 358 กฎการตัดสินใจจากทั้งหมด 84 สาขาวิชา โดยสามารถนำกฎการตัดสินใจที่ได้มาพัฒนาเป็นระบบสนับสนุนการตัดสินใจเลือกสาขาวิชาในระบบ TCAS มหาวิทยาลัยมหาสารคาม เพื่อให้ผู้สมัครได้ทดสอบการเลือกสาขาวิชาก่อนการเลือกสาขาวิชาจริง เพื่อประเมินโอกาสในการผ่านเข้าศึกษา

ข้อเสนอแนะ

งานวิจัยนี้ใช้ข้อมูลในการวิเคราะห์ คือข้อมูลผลคะแนนการทดสอบความรู้ทั่วไป (GAT) และผลคะแนนการทดสอบความรู้ทางวิชาการและวิชาชีพ (PAT) โดยเป็นคะแนนที่ใช้ในการพิจารณาผลผู้ผ่านการคัดเลือกและมีสิทธิ์ในการยืนยันสิทธิ์ในระบบ TCAS โดยการรับสมัครคัดเลือกในระบบ TCAS นั้นมีการสมัคร 5 รอบ ซึ่งรอบที่ 1 นั้นการพิจารณาผลคัดเลือกไม่ได้ใช้คะแนน GAT และคะแนน PAT ซึ่งพิจารณาผลด้วยแฟ้มสะสมผลงาน (Portfolio) ซึ่งในอนาคตผู้วิจัยมีแนวคิดที่จะนำปัจจัยด้านแฟ้มสะสมผลงานมาประกอบการวิเคราะห์เพื่อให้ครบคุมการรับสมัครคัดเลือกในระบบ TCAS ในทุกๆ รอบการรับสมัคร

กิตติกรรมประกาศ

โครงการวิจัยนี้ได้รับการสนับสนุนจากเงินทุนอุดหนุนการวิจัยจากงบประมาณเงินรายได้ ประจำปี 2563 มหาวิทยาลัยมหาสารคาม

เอกสารอ้างอิง

กองบริการการศึกษา มหาวิทยาลัยมหาสารคาม. (2563). *ระเบียบการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระดับปริญญาตรี*. ค้นเมื่อวันที่ 8 มกราคม 2563 จาก https://admission.msu.ac.th/

กองบริการการศึกษา มหาวิทยาลัยมหาสารคาม. (2563). *สรุปผลการดำเนินงานการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระดับปริญญาตรี*. กองบริการการศึกษา มหาวิทยาลัยมหาสารคาม.

ชิดชนก ส่งศิริ, ธนาวินท์ รักธรรมานนท์ และกฤษณะ ไวยมัย. (2544). การใช้เทคนิค Data Mining เพื่อค้นหาภาควิชาที่เหมาะสมที่สุดให้กับนิสิต. *การประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ สาขาวิศวกรรมศาสตร์ ครั้งที่ 39* ; วันที่ 5-7 กุมภาพันธ์ 2544. กรุงเทพมหานคร: 43-50.

ชุติมา อุตมะมุณีย์ และประสงค์ ปราณีตพลกรัง. (2553). *การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาออนไลน์สำหรับเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา*. ค้นเมื่อวันที่ 20 มกราคม 2563 จาก https://ph02.tci-thaijo.org/index.php/JIST/article/download/135149/101010/

ฉัตรเกล้า เจริญผล. (2556). *เอกสารประกอบการสอนรายวิชา Introduction to Data Mining 2013*. คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม.

สุพัฒน์กุล ภัคโชค (2556). ตัวแบบการเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลาย โดยการพิจารณาผลการเรียนรายวิชาหลัก ด้วยเทคนิคเหมืองข้อมูล. (วิทยานิพนธ์ปริญญาวิทยาศาสตร์มหาบัณฑิต). กรุงเทพมหานคร: มหาวิทยาลัยธุรกิจบัณฑิตย์.

สมาคมที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.). (2561) *ประกาศกระทรวงศึกษาธิการ นโยบายการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่.*

สถาบันทดสอบทางการศึกษาแห่งชาติ องค์การมหาชน (สทศ.). (2563). *ระบบรายงานผลการสอบเพื่อการคัดเลือกบุคคลเข้าศึกษาระดับอุดมศึกษา TCAS*. ค้นเมื่อวันที่ 15 มกราคม 2563 จาก http://escorereport.niets.or.th/

อนันต์ ปินะเต. (2559). การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสมัครในสาขาวิชาโดยใช้เทคนิคต้นไม้ตัดสินใจ. *วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยมหาสารคาม*, *35*(4), 411-421.

Hand, J., and Kamber, M. (2001). *Data Mining Concepts and Techniques*. Retrieved January 10, 2020, from https://cs.wmich.edu/~yang/teach/cs595/han/ch01.pdf

Markov, Z. *An Introduction to the WEKA Data Mining System.* January 10, 2020, https://cs.ccsu.edu/~markov/weka-tutorial.pdf