

การประยุกต์เหมืองข้อมูลในการศึกษาปัจจัยในการเลือกเข้าศึกษาต่อในมหาวิทยาลัยผ่านการ ทำแบบจำลองการจัดกลุ่ม

Application of Data Mining in Studying the Selection of University Admission through Clustering Modeling

ประจักษ์ เชิดโฉม¹

Prajak Chertchom

บทคัดย่อ (Abstract)

บทความวิจัยนี้เป็นการประยุกต์ใช้การทำเหมืองข้อมูลและเทคนิค K-means Clustering ในการวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัย ผ่านการจัดกลุ่มนักศึกษาตามลักษณะอาทิเช่น คะแนนเฉลี่ยของมัธยมปลาย ภูมิลำเนา เป็นต้น เพื่อพัฒนาแบบจำลองการจัดกลุ่มในการวิเคราะห์ปัจจัยการเลือกเข้าศึกษาของคณะเศรษฐศาสตร์และบริหารธุรกิจในมหาวิทยาลัยแห่งหนึ่งในภาคใต้ และตรวจสอบความเหมาะสมและประสิทธิภาพของแบบจำลอง จากข้อมูล Clustering ที่ได้จากการวิเคราะห์ จากการแบ่งกลุ่มนักศึกษาของคณะเศรษฐศาสตร์และบริหารธุรกิจเป็น 3 กลุ่ม นักศึกษาที่เลือกเรียนในภาคปีการศึกษาตั้งแต่ 2562-2564 จำนวน 2,238 คน จากการทำ Clustering ด้วย K-means แล้ว พบว่ากลุ่มที่ 1 มาจากจังหวัดสงขลามากที่สุด โดยมีจำนวนข้อมูล 1,379 คน และมีค่าเฉลี่ยของ GPA อยู่ที่ 3.2625 ดังนั้น ฐานลูกค้าของมหาวิทยาลัย คือ นักเรียนที่อยู่ในจังหวัดสงขลาที่เกรดเฉลี่ยอยู่ในระดับ “High” คือ 3.00-3.50 สำหรับการวัดความแม่นยำ (Accuracy) ใน Clustering DB Index จากโมเดลได้เท่ากับ -0.680 ในกรณีนี้ถือว่าใกล้เคียง 0 แสดงว่าการแบ่งกลุ่มเป็นอย่างมีประสิทธิภาพ ซึ่งผลการวิเคราะห์จะช่วยให้สถาบันการศึกษาสามารถพัฒนานโยบายและกิจกรรมที่เกี่ยวข้องให้เหมาะสมกับการรับสมัครนักศึกษาในภาวะการแข่งขันที่รุนแรงในปัจจุบัน

คำสำคัญ (Keywords) : การทำเหมืองข้อมูล; แบบจำลองการจัดกลุ่ม; การคัดเลือกเข้าศึกษาต่อ

Abstract

In this research paper data mining and K-means clustering techniques were employed to examine the factors influencing the selection process for university admissions. The study involved categorizing students based on their characteristics, such as their high school average score and domicile, to develop a clustering model that could analyze the factors impacting the selection of studies in the Faculty of Economics and Business Administration at a university located in the South. The effectiveness and suitability of the model were tested using clustering data derived from the analysis. The results of the study indicate that the students in the Faculty of Economics and Business Administration could be grouped into three distinct

Received: 2023-04-30 Revised: 2023-05-16 Accepted: 2023-05-17

¹ อาจารย์ประจำหลักสูตรวิชาการค้าสมัยใหม่และนวัตกรรมบริการ คณะเศรษฐศาสตร์และบริหารธุรกิจ มหาวิทยาลัยทักษิณ Lecture of Program in Modern Trade and Service Innovation : B.B.A. (Modern Trade and Service Innovation), Faculty of Economics and Business Administration, Thaksin University. E-mail: prajak.c@tsu.ac.th

categories. Of the 2,238 students who opted to study in the academic year 2019-2021, the first group had the most students from Songkhla province, with 1,379 individuals having an average GPA of 3.2625. As a result, the university's customer base appears to be comprised of students from Songkhla province who have a high-grade point average, which is between 3.00-3.50 as per the measurement of accuracy (Accuracy) in clustering. The DB Index from the model is -0.680, which indicates that the segmentation is effective since it is close to 0. These findings can help educational institutions in developing policies and activities that are more relevant to student recruitment, especially in the context of the current competitive landscape.

Keywords: Data Mining; Clustering Model; University admission selection

บทนำ (Introduction)

การพัฒนาการทำการตลาดเพื่อหานักศึกษาในสถาบันการศึกษาและลดต้นทุนในการจัดกิจกรรมประชาสัมพันธ์ เป็นเรื่องที่ทำนายในยุคปัจจุบันสำหรับหลายสถาบัน การทำวิจัยนี้มีวัตถุประสงค์เพื่อประยุกต์ใช้เทคนิคการเหมืองข้อมูลในการวิเคราะห์ข้อมูลปัจจัยที่ส่งผลต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัยของนักศึกษา โดยใช้ข้อมูลจากระบบทะเบียนนักศึกษาจากมหาวิทยาลัยรัฐแห่งหนึ่งในภาคใต้ ตั้งแต่ปีการศึกษา 2562 ถึง 2564 และนำเทคนิคการแบ่งกลุ่มด้วยวิธี K-Means Clustering มาวิเคราะห์ข้อมูล เพื่อหากลุ่มของปัจจัยที่มีผลต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัยในปี 2564 สถานการณ์ของอุดมศึกษาไทยแสดงให้เห็นถึงความจำเป็นของการปรับตัวของมหาวิทยาลัยในการรับนักศึกษามากยิ่งขึ้น เนื่องจากจำนวนนักศึกษาที่รับสมัครผ่านระบบ TCAS 61-64 ในแต่ละรอบมีจำนวนน้อยกว่าที่มหาวิทยาลัยสามารถรับได้กล่าวคือข้อมูลสถิติเกี่ยวกับการสอบ GAT/PAT ในระบบ TCAS ในช่วงปีการศึกษา 2561-2564 ปี 2561 (GAT/PAT62) มีจำนวนผู้สมัครสอบทั้งหมด 265,552 คน และจำนวนผู้สมัครสอบน้อยที่สุดเกิดขึ้นในปี 2562 (GAT/PAT63) ที่มีจำนวนผู้สมัครสอบ 255,295 คน โดยในปีการศึกษา 2564 มีผู้สมัครสอบ จำนวน 257,282 คน โดยที่จำนวนมหาวิทยาลัย 155 ทั้งมหาวิทยาลัยรัฐและเอกชน แห่งในประเทศไทย สามารถรับนักศึกษาใหม่ได้ประมาณ 420,000 ที่นั่งต่อปี แต่จากสถิติการสอบผู้สมัครสอบ TGAT/TPAT64 มีเพียงประมาณ 257,282 คนเท่านั้นและยังมีการสละสิทธิ์อย่างมากมาย หากภาวะนี้ยังคงเกิดขึ้นต่อไป มหาวิทยาลัยอาจเจอกับวิกฤตในอนาคต ดังนั้น การวิเคราะห์สถานการณ์เป็นสิ่งที่สำคัญและจำเป็นต่อการปรับปรุงและพัฒนาการศึกษาของมหาวิทยาลัยในอนาคต หากมหาวิทยาลัยไม่ปรับตัวตามสถานการณ์นี้ ก็อาจเสียเปรียบไม่น้อยเช่นกัน ในกรณีที่มหาวิทยาลัยไม่สามารถรับมือกับภาระขาดทุนได้ อาจต้องปิดไป เป็นต้น (สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม, 2566)

การใช้เทคนิค K-means Clustering เป็นเทคนิคการแบ่งกลุ่มข้อมูล (clustering) ที่ใช้วิธีการแบ่งกลุ่มโดยทำการคำนวณระยะห่างของข้อมูล โดยการจัดกลุ่มที่มีค่าความคล้ายคลึงกันสูงสุดในกลุ่มเดียวกัน โดยแบ่งข้อมูลเป็น K กลุ่ม โดย K คือจำนวนกลุ่มที่ต้องการจัดแบ่งโมเดล K-means Clustering ใช้งานได้กับการจัดกลุ่มข้อมูลในหลายๆ สาขา (King, R. S., 2015: 5) เช่น การจัดกลุ่มหมู่เสี่ยงโรค หรือการจัดกลุ่มข้อมูลทางการศึกษา เป็นต้น โดยจุดเด่นของเทคนิค K-means Clustering ในการวิเคราะห์ข้อมูลทางธุรกิจที่มีจำนวนมากคือ มีความเร็วในการทำงานสูง ดังนั้นถ้ามีจำนวนข้อมูลที่มากมาย ก็ยังสามารถจัดกลุ่มข้อมูลได้อย่าง

รวดเร็ว มีผลลัพธ์ที่สามารถตีความได้ง่าย (Asif, R. et al., 2017: 177: Goyal, M., & Vohra, R., 2012: 118) โดยเทคนิค K-means Clustering สามารถนำไปใช้ในธุรกิจได้หลากหลายแบบ ดังต่อไปนี้ เช่น การจัดกลุ่มลูกค้า ตัวอย่างเช่น องค์กรธุรกิจสามารถใช้ เทคนิค K-means Clustering เพื่อจัดกลุ่มลูกค้าเพื่อให้เข้าใจลักษณะและพฤติกรรม การซื้อของลูกค้าได้ง่ายขึ้น (Papamitsiou, Z., & Economides, A. A., 2014: 51)

วัตถุประสงค์การวิจัย (Research Objectives)

การศึกษานี้ได้ทำการใช้เทคนิค K-means Clustering โดยมีวัตถุประสงค์การวิจัย ดังนี้

1. เพื่อแบ่งกลุ่มนักศึกษาตามปัจจัยในการเลือกเข้าศึกษาต่อในมหาวิทยาลัยที่มีต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัย
2. เพื่อพัฒนาแบบจำลองการจัดกลุ่มเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัย ทดสอบและเปรียบเทียบแบบจำลองการจัดกลุ่มด้วยข้อมูลจริงเพื่อตรวจสอบความเหมาะสมและประสิทธิภาพของแบบจำลองจากเครื่องมือวิเคราะห์เหมืองข้อมูล WEKA และ RapidMiner
3. เพื่อพัฒนานโยบายและกิจกรรมที่เกี่ยวข้องให้เหมาะสมกับการรับสมัครนักศึกษาในภาวะการแข่งขันที่รุนแรงในปัจจุบัน

วิธีดำเนินการวิจัย (Research Methods)

งานวิจัยครั้งนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) เป็นการวิจัยที่ค้นหาความรู้จากข้อมูลขนาดใหญ่โดยใช้เทคนิคเหมืองข้อมูล (Data Mining) แบบจำลองการจัดกลุ่ม (K-Mean Clustering) ซึ่งจะนำผลการวิจัยมาใช้ในการวางแผนการรับนักศึกษาระดับปริญญาตรี

1. ขอบเขตของการวิจัย

1.1 ขอบเขตด้านเนื้อหา

การวิจัยเพื่อหาคำตอบว่าปัจจัยใดที่ส่งผลต่อการตัดสินใจของนักเรียนที่จะเลือกเข้าศึกษาต่อในมหาวิทยาลัย โดยใช้ข้อมูลที่ได้รับจากระบบทะเบียนนักศึกษา โดยจะเน้นการวิเคราะห์การจัดกลุ่มระหว่างตัวแปรที่เกี่ยวข้องกัน เช่น คะแนนผลการเรียนเฉลี่ยของนักเรียนในระดับชั้นมัธยมศึกษาตอนปลาย จังหวัดที่อาศัยและสถานภาพของครอบครัว

1.2 ขอบเขตด้านประชากร

การเก็บรวบรวมข้อมูลในการศึกษาครั้งนี้ คือ การนำข้อมูลนักศึกษาจากระบบทะเบียนของมหาวิทยาลัยรัฐแห่งหนึ่งในจังหวัดสงขลา ระหว่าง ปีการศึกษา 2562 ถึงปีการศึกษา 2564 โดยเลือกข้อมูลเฉพาะคณะเศรษฐศาสตร์และบริหารธุรกิจจำนวน 2,238 คน (รายการ)

1.3 ขอบเขตด้านพื้นที่

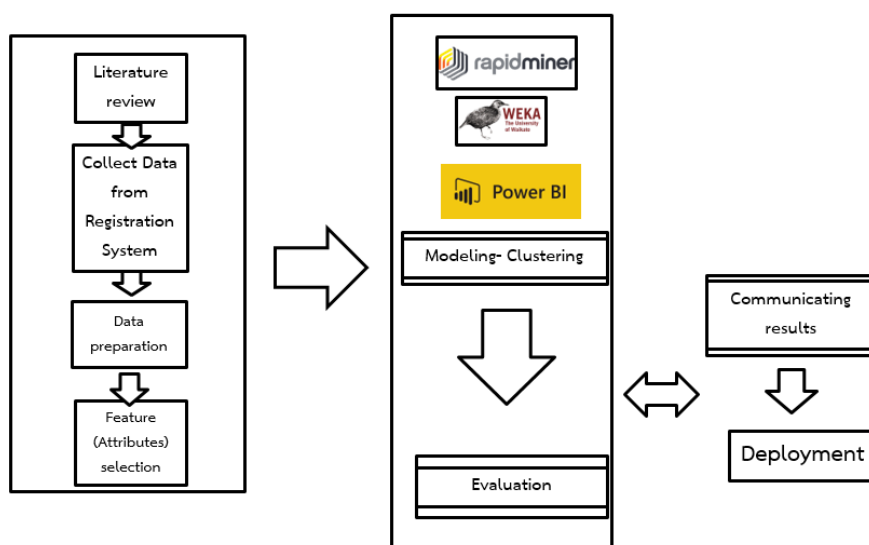
งานวิจัยนี้ได้ทำการวิจัยจากข้อมูลระบบทะเบียนของมหาวิทยาลัยแห่งหนึ่งในจังหวัดสงขลา

1.4 ขอบเขตด้านระยะเวลา

งานวิจัยนี้ได้ทำการวิจัยจากข้อมูลระบบทะเบียนของมหาวิทยาลัยแห่งหนึ่งในจังหวัดสงขลา จากระบบทะเบียนของมหาวิทยาลัยรัฐแห่งหนึ่งในภาคใต้ ของปีการศึกษา 2562 ถึงปีการศึกษา 2564

2. เครื่องมือที่ใช้ในการวิจัย

ผู้วิจัยได้เลือกเทคนิคและโมเดลในการทำเหมืองข้อมูลด้วยเทคนิคการแบ่งกลุ่ม (Clustering) ซึ่งเป็นวิธีการจัดกลุ่มข้อมูลเพื่อแบ่งข้อมูลออกเป็นกลุ่มย่อยๆ ที่มีลักษณะคล้ายคลึงกัน โดยที่ไม่ต้องมีความรู้ล่วงหน้าเกี่ยวกับกลุ่มที่อยู่ภายในข้อมูลมาทำการศึกษาและคัดข้อมูลที่มีลักษณะเดียวกันมาอยู่ในกลุ่มเดียวกันโดยการใช้ K-Means Algorithm จากสมการ $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ (Celebi, M. E. (Ed.), 2014: 147) นอกจากนี้ผู้วิจัยนำกระบวนการทำเหมืองข้อมูลในรูปแบบ (Cross Industry Standard Process for Data Mining: CRISP-DM) [10] ดังรูปที่ 1 กรอบการวิจัย (Research Framework) มาประกอบการดำเนินงาน (Martínez-Plumed, F et al., 2019: 3048)



รูปที่ 1 กรอบการวิจัย (Research Framework)

2.1 สำหรับเครื่องมือที่ใช้ในการทำวิจัย ผู้วิจัยใช้เครื่องมือ 3 โปรแกรมคือ Power BI ซึ่งเป็นเครื่องมือที่ถูกใช้ในการวิเคราะห์และสร้างรายงานข้อมูลเบื้องต้นเชิงพรรณนา (Descriptive) เพื่อสร้าง Visualization และ Dashboard เครื่องมือที่สองคือ Weka เป็นเครื่องมือซอฟต์แวร์โอเพนซอร์ส (open-source software) ที่ใช้สำหรับการทำงานทางด้าน machine learning และ data mining ผู้วิจัยใช้เพื่อช่วยในการวิเคราะห์ข้อมูลแบบตรงเทียบกับ Rapid Miner ซึ่งเป็นเครื่องมือที่สาม และทำการเปรียบเทียบผลจากสองโปรแกรมด้านการทำเหมืองข้อมูลระหว่าง WEKA และ RapidMiner

2.2 จากรูปที่ 1 ผู้วิจัยเริ่มกระบวนการเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนสำคัญในการวิเคราะห์ข้อมูล ซึ่งมีวัตถุประสงค์เพื่อเตรียมข้อมูลให้เหมาะสมกับการนำไปวิเคราะห์ โดยประกอบด้วยขั้นตอนดังนี้ (Ayele, W. Y., 2020: 20) พร้อมทั้งทำการทบทวนวรรณกรรม

การเก็บข้อมูล (Data Collection) จากระบบทะเบียน

การเตรียมข้อมูล (Data Preparation) และตรวจสอบข้อมูล (Data Inspection) คือการเตรียมข้อมูลก่อนนำเข้าสู่กระบวนการวิเคราะห์หรือการประมวลผลข้อมูลต่าง ๆ เพื่อให้ข้อมูลมีคุณภาพและเหมาะสมสำหรับการวิเคราะห์ โดยรวมถึงการเลือกข้อมูลที่จะนำเข้าสู่กระบวนการวิเคราะห์ ตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล แปลงรูปแบบของข้อมูลเป็นต้น

ตารางที่ 1 การแปลงข้อมูล GPA (Grade Point Average)

Range Of GPA	Scale	Number
3.50-4.00	Excellent	5
3.00-3.49	High	4
2.75-2.99	Good	3
2.50-2.74	medium	2
2.00-2.49	Poor	1
1.50-1.99	Bad	0
N/A	(blank)	7

จากตารางที่ 1 ผู้วิจัยได้ทำการแปลงข้อมูล 2 ครั้ง โดยนำเกรดจากระบบทะเบียนรายบุคคลมาจัดระดับ 7 ระดับดังตารางและใส่หมายเลขในแต่ละระดับ

ตารางที่ 2 ตัวอย่างการแปลงข้อมูลจังหวัด (Province)

Province	Province Index No.
Bangkok	1
Buriram	2
Chiang Rai	3
Chonburi	4
Narathiwat	11
Nonthaburi	12
Pathum Thani	13
Pattani	14
Songkhla	25
Surat Thani	26
Surin	27
Trang	28
Udon Thani	29
Yala	30
(blank)	31

จากตารางที่ 1 และ 2 คือการทำความสะอาดข้อมูล (Data Cleaning): เป็นการลบข้อมูลที่ไม่เหมาะสมหรือไม่เกี่ยวข้องออก ผู้วิจัยได้ทำการแปลงข้อมูล 2 ครั้ง โดยนำจังหวัดที่นักศึกษาอาศัยมาแปลงเป็นภาษาอังกฤษและตัวเลข เพื่อให้ง่ายต่อการใช้ Software ต่างประเทศในการทำภาพเสมือน (Visualization) และให้ง่ายสำหรับการประมวลผล

การเลือกข้อมูล (Data Selection): เป็นการเลือกข้อมูลที่เกี่ยวข้องกับการวิเคราะห์ เพื่อลดเวลาในการวิเคราะห์

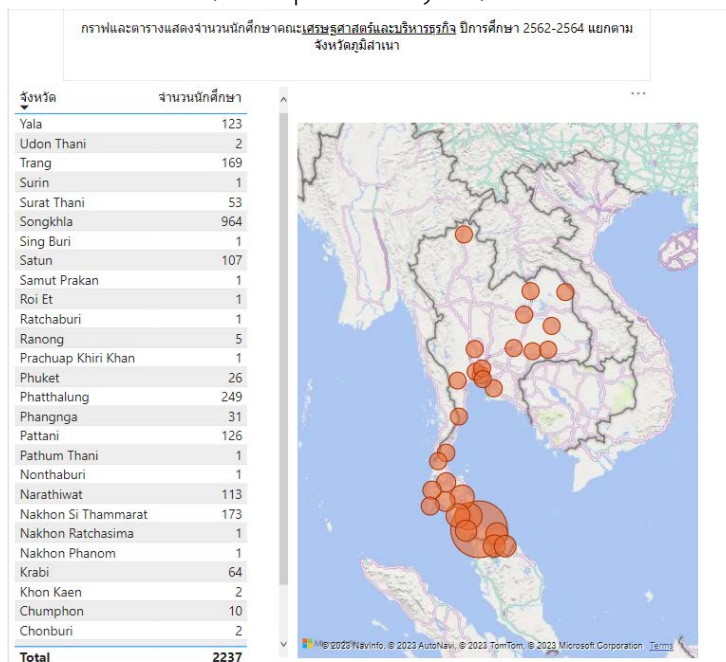
การเตรียมข้อมูล (Data Transformation): เป็นการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับการนำไปวิเคราะห์ โดยการแปลงข้อมูล เป็นตัวเลข (Numerical Feature) เพื่อให้สามารถวิเคราะห์ด้วยเทคนิค K-means Clustering ได้ดังตารางที่ 1 และตารางที่ 2

3. การวิเคราะห์ข้อมูลเป็นการวิเคราะห์ข้อมูลด้วยเทคนิค K-means ซึ่งเป็นวิธีการแบ่งกลุ่มข้อมูล (Clustering) ที่ใช้ความสัมพันธ์ระหว่างข้อมูลเพื่อแบ่งข้อมูลออกเป็นกลุ่มตามความคล้ายคลึงกัน โดยจะใช้

ข้อมูลตัวเลข หรือคุณลักษณะของข้อมูลเพื่อทำการคำนวณหาจุดศูนย์กลางของกลุ่ม (Centroid) แล้วจัดกลุ่มข้อมูลที่มีค่าใกล้เคียงกับจุดศูนย์กลางเดียวกัน แล้วทำการวิเคราะห์และอธิบายคุณสมบัติของแต่ละกลุ่ม

ผลการวิจัย (Research Result)

ผลการวิเคราะห์เชิงพรรณนา (Descriptive Analytics)



รูปที่ 2 กราฟและตารางแสดงจำนวนนักศึกษาคณะเศรษฐศาสตร์และบริหารธุรกิจ ปีการศึกษา 2562-2564 แยกตามจังหวัดอุบลราชธานี

รูปที่ 2 เป็นกราฟและตารางที่แสดงจำนวนนักศึกษาคณะเศรษฐศาสตร์และบริหารธุรกิจ ปีการศึกษา 2562-2564 แยกตามจังหวัดอุบลราชธานีจัดทำโดย POWER BI โดยแสดงจำนวนนักศึกษาที่เรียนอยู่ในคณะเศรษฐศาสตร์และบริหารธุรกิจในแต่ละจังหวัดของประเทศไทย กราฟแสดงข้อมูลในรูปแบบแผนที่ภูมิภาคของประเทศไทย โดยแต่ละจังหวัดถูกแทนด้วยสัญลักษณ์เป็นวงกลมขนาดต่าง ๆ ซึ่งขนาดของวงกลมแสดงถึงจำนวนนักศึกษาในแต่ละจังหวัด จากกราฟจะเห็นได้ว่ามีจำนวนนักศึกษาในคณะเศรษฐศาสตร์และบริหารธุรกิจที่เรียนอยู่มากที่สุดอยู่ในจังหวัดสงขลา ตามด้วยจังหวัดภูเก็ตและภาคใต้ของประเทศไทย

ตารางแสดงจำนวนนักศึกษาที่เรียนอยู่ในคณะเศรษฐศาสตร์และบริหารธุรกิจในแต่ละจังหวัด แสดงจำนวนนักศึกษาในแต่ละจังหวัดของประเทศไทย โดยจังหวัดที่มีนักศึกษาเรียนอยู่มากที่สุด 5 ลำดับคือ สงขลา มีนักศึกษาเรียนอยู่ 964 คน ตามมาด้วยจังหวัดพัทลุง โดยมีจำนวนนักศึกษา 249 คน จังหวัด นครศรีธรรมราชมีจำนวนนักศึกษา 173 คน จังหวัดตรัง นครศรีธรรมราชมีจำนวนนักศึกษา 169 คน จังหวัดปัตตานีมีจำนวนนักศึกษา 126 คน และจังหวัดยะลามีจำนวนนักศึกษา 123 คนตามลำดับ

Final cluster centroids:

Attribute	Full Data (2237.0)	Cluster#		
		0 (1379.0)	1 (411.0)	2 (447.0)
BEFOREGPA-Num	3.2682	3.2625	2.0803	4.3781
PROVINCE_Num	20.9428	25.7542	12.9367	13.4609

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	1379 (62%)
1	411 (18%)
2	447 (20%)

รูปที่ 3 การทำโมเดล Clustering ด้วย WEKA

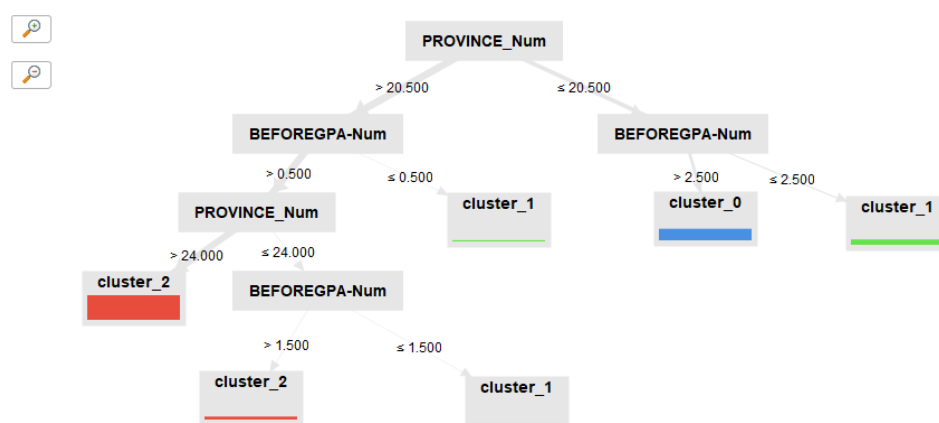
จากรูปที่ 3 แสดงตาราง Clustering ซึ่งเป็นผลการทำ Clustering บนข้อมูล 2 ตัวแปรคือ เกรดมัธยมปลาย (GPA) (BEFOREGPA-Num) และ จังหวัดภูมิลำเนา (PROVINCE_Num) โดยจำนวนข้อมูลทั้งหมดคือ 2,237 ตัวอย่าง ซึ่งแบ่งเป็น 3 กลุ่ม กลุ่มที่ 0, 1 และ 2 โดยแต่ละกลุ่มจะมีจำนวนข้อมูลและค่าเฉลี่ยของแต่ละตัวแปร ดังนี้

กลุ่มที่ 0 มีจำนวนข้อมูล 1,379 ตัว โดยมีค่าเฉลี่ยของ BEFOREGPA-Num เท่ากับ 3.2625 และมีค่าเฉลี่ยของ PROVINCE_Num เท่ากับ 25.7542

กลุ่มที่ 1 มีจำนวนข้อมูล 411 ตัว โดยมีค่าเฉลี่ยของ BEFOREGPA-Num เท่ากับ 2.0803 และมีค่าเฉลี่ยของ PROVINCE_Num เท่ากับ 12.9367

กลุ่มที่ 2 มีจำนวนข้อมูล 447 ตัว โดยมีค่าเฉลี่ยของ BEFOREGPA-Num เท่ากับ 4.3781 และมีค่าเฉลี่ยของ PROVINCE_Num เท่ากับ 13.4609 ซึ่งหมายความว่าข้อมูลในกลุ่มนี้มีค่าเฉลี่ยของ BEFOREGPA-Num และ PROVINCE_Num สูงกว่ากลุ่ม 0 และกลุ่มที่ 1

k-Means - Cluster Tree



รูปที่ 4 การทำโมเดล K-means Cluster Tree ด้วย RapidMiner

จาก Cluster Tree ที่ได้จากการใช้เทคนิค K-means Clustering ด้วย RapidMiner จะสามารถอธิบายได้ ดังนี้

มีทั้งหมด 3 กลุ่มหลักๆ โดยแต่ละกลุ่มจะแบ่งออกเป็นกลุ่มย่อยๆ ไปอีก โดยกลุ่มหลักแรก (Cluster 0) มีจำนวนข้อมูลน้อยที่สุด และมีค่าเฉลี่ยของเกรด GPA ระดับมัธยมปลาย (BEFOREGPA-Num) และมีค่าเฉลี่ย BEFOREGPA-Num สูงสุด ที่ 3.955 กลุ่มหลักที่สอง (Cluster 1) มีจำนวนข้อมูลที่น้อยที่สุด และมีค่าเฉลี่ย BEFOREGPA-Num ต่ำกว่ากลุ่มอื่นๆ ที่ 1.470 กลุ่มหลักที่สาม (Cluster 2) มีจำนวนข้อมูลมากที่สุด มีค่าเฉลี่ยของรหัสจังหวัด (PROVINCE_Num) สูงสุด ที่ 25.692 และมีค่าเฉลี่ย BEFOREGPA-Num อยู่ที่ 3.301

สำหรับการวัดความแม่นยำ (Accuracy) เป็นการวัดความเหมือนกันของข้อมูลที่อยู่ใน cluster เดียวกันว่ามีลักษณะคล้ายคลึงกันมากแค่ไหน โดยใช้ตัววัดที่เรียกว่า Davies Bouldin เป็น DB Index เป็นตัววัดความคลาดเคลื่อนในการแบ่งกลุ่มของ Clustering Algorithm ที่คำนวณด้วยความสัมพันธ์ระหว่างระยะทางระหว่าง Centroids ของแต่ละกลุ่มและขนาดของกลุ่ม เป็นการวัดว่าแต่ละกลุ่ม (Cluster) แต่ละกลุ่มมีความแตกต่างกันมากน้อยเพียงใด ใน DB Index ค่าที่น้อยกว่าจะแสดงถึงความแตกต่างของกลุ่ม (Cluster) ที่น้อยกว่า และในกรณีที่ DB Index เป็น 0 จะแสดงว่าการแบ่งกลุ่มเป็นอย่างมีประสิทธิภาพ ซึ่งจากโมเดลได้เท่ากับ -0.680 ในกรณีนี้ถือว่าใกล้เคียง 0 แสดงว่าการแบ่งกลุ่มเป็นอย่างมีประสิทธิภาพ

อภิปรายผลการวิจัย (Research Discussion)

จากข้อมูล Clustering ที่ได้จากการวิเคราะห์ พบว่ามีการแบ่งกลุ่มนักศึกษาของคณะเศรษฐศาสตร์และบริหารธุรกิจเป็น 3 กลุ่ม โดยจำนวนข้อมูลแต่ละกลุ่มมีดังนี้

Cluster 0: มีนักศึกษาอยู่ในกลุ่ม (Cluster) ที่ 1 จำนวน 1,379 คน

Cluster 1: มีนักศึกษาอยู่ในกลุ่ม (Cluster) ที่ 2 จำนวน 411 คน

Cluster 2: มีนักศึกษาอยู่ในกลุ่ม (Cluster) ที่ 3 จำนวน 447 คน

จากการดูค่า centroid ของแต่ละกลุ่ม พบว่ามีความแตกต่างกันอย่างชัดเจน โดยเฉลี่ยของ ระดับเกรดมัธยมปลาย (BEFOREGPA-Num) ซึ่งเทียบกับตารางที่ 2 (มี 7 ระดับ) แต่ละกลุ่มมีค่าดังนี้

Cluster 0: 3.2625 ค่าเฉลี่ยของระดับเกรดมัธยมปลาย (BEFOREGPA-Num) ของกลุ่ม (Cluster) ที่ 0 เท่ากับ 3.2625

Cluster 1: 2.0803 ค่าเฉลี่ยของระดับเกรดมัธยมปลาย (BEFOREGPA-Num) ของ กลุ่ม (Cluster) ที่ 1 เท่ากับ 2.0803

Cluster 2: 4.3781 ค่าเฉลี่ยของระดับเกรดมัธยมปลาย (BEFOREGPA-Num) ของ กลุ่ม (Cluster) ที่ 2 เท่ากับ 3.2625

และเฉลี่ยของรหัสจังหวัด (PROVINCE_Num) แต่ละกลุ่มมีค่าดังนี้

Cluster 0: 25.7542, Cluster 1: 12.9367, Cluster 2: 13.4609

สรุปได้ว่าจากวัตถุประสงค์ข้อที่ 1 แบ่งกลุ่มนักศึกษาตามปัจจัยในการเลือกเข้าศึกษาต่อในมหาวิทยาลัยที่มีต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัย จากการทดลองผู้วิจัยได้แบ่งกลุ่ม (Cluster) ออกเป็น 3 กลุ่ม ผลวิจัยพบว่านักศึกษาที่เลือกเรียนคณะเศรษฐศาสตร์และบริหารธุรกิจ ปีการศึกษา 2562-

2564 กลุ่มที่มีจำนวนข้อมูลที่มากที่สุดอยู่ในกลุ่ม 0 ซึ่งเป็นกลุ่มนักเรียนจากจังหวัดสงขลา โดยมีจำนวนข้อมูลมากถึง 1,379 คน ส่วนในกลุ่มนี้มีค่าเฉลี่ยของ GPA อยู่ที่ 3.2625 ซึ่งใกล้เคียงกับค่าเฉลี่ยของ GPA ทั้งหมดที่อยู่ระดับ 3.2682 แปลว่านักศึกษาส่วนใหญ่มาจากจังหวัดสงขลา ดังนั้น ฐานลูกค้าของมหาวิทยาลัย คือนักเรียนที่อยู่ในจังหวัดสงขลา ที่เกรดเฉลี่ยอยู่ในระดับ “High” คือ 3.00-3.50 ซึ่งสอดคล้องกับผลการวิจัยของ Goyal, M., & Vohra, R., (2012: 113) ที่มีการเสนอให้ใช้เทคนิคเหมืองข้อมูลเพื่อเพิ่มประสิทธิภาพของสถาบันการศึกษา โดยใช้เทคนิคเหมืองข้อมูล เช่น การจัดกลุ่มข้อมูล (clustering) ต้นไม้การตัดสินใจ (decision tree) และสมมติฐาน (association) ไปใช้กับกระบวนการทางการศึกษาของสถาบันการศึกษา ซึ่งจะช่วยให้เพิ่มประสิทธิภาพของการเรียนการสอนของนักเรียนได้

สำหรับวัตถุประสงค์ที่ 2 พัฒนาแบบจำลองการจัดกลุ่มเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกเข้าศึกษาต่อในมหาวิทยาลัย ทดสอบและเปรียบเทียบแบบจำลองการจัดกลุ่มด้วยข้อมูล จากรูปที่ 4 พบว่า ผลการทดสอบและเปรียบเทียบแบบจำลองจัดกลุ่มแสดงให้เห็นว่าแบบจำลองที่พัฒนาด้วย RapidMiner ให้ผลลัพธ์ที่เหมือนกับ WEKA และปัจจัยที่ส่งผลต่อการเลือกมหาวิทยาลัยแห่งนี้คือใกล้ภูมิลำเนา นั่นก็คือจังหวัดสงขลา และจังหวัดใกล้เคียง เช่น ปัตตานี นราธิวาส และพัทลุง เป็นต้น โดยระดับคะแนนของนักศึกษาที่สมัครมาอยู่ในระดับ 3.00-3.50 DB Index ซึ่งจากโมเดลได้เท่ากับ -0.680 ในกรณีนี้ถือว่าใกล้เคียง 0 แสดงว่าการแบ่งกลุ่มเป็นอย่างมีประสิทธิภาพ ซึ่งสอดคล้องกับ Larose, D. T., & Larose, C. D. (2014: 149) และ Davies, D. L., & Bouldin, D. W. (1979: 224-227) ที่อธิบายถึงการวัดเป็นระยะทางคลัสเตอร์ขั้นต่ำที่ยอมรับได้หรือค่าเบี่ยงเบนมาตรฐานขั้นต่ำที่ยอมรับได้

สำหรับวัตถุประสงค์ที่ 3 นำผลลัพธ์ที่ได้มาวิเคราะห์และสรุปผลเพื่อให้สถาบันการศึกษาสามารถพัฒนานโยบายและกิจกรรมที่เกี่ยวข้องให้เหมาะสมกับการรับสมัครนักศึกษาในภาวะการแข่งขันที่รุนแรงในปัจจุบันนี้ ผู้วิจัยเสนอว่าการทำการตลาดเพื่อหานักศึกษาสามารถทำได้หลากหลายวิธีโดยสอดคล้องกับงานวิจัยของ Papamitsiou, Z., & Economides, A. A. (2014: 49) และ Goyal, M., & Vohra, R. (2012: 113) ต่อไปนี้ คือ วิธีการทำการตลาดที่อาจช่วยให้สถาบันการศึกษาได้รับความสนใจจากนักศึกษาที่เป็นเป้าหมาย:

1. การสร้างกิจกรรมทำความรู้จักสถาบันการศึกษาผ่านโฆษณา เช่น จัดนิทรรศการ งานแสดงสินค้า หรือกิจกรรมกีฬา การโฆษณาบนสื่อออนไลน์ เช่น โซเชียลมีเดีย เป็นต้น เพื่อเพิ่มโอกาสในการเห็นและทำความรู้จักสถาบันการศึกษาในเขตพื้นที่ที่ใกล้กับที่ตั้งของมหาวิทยาลัยเป็นหลัก เพราะนักศึกษาในรอบ 4 ปีที่ผ่านมาจากจังหวัดใกล้เคียง และไม่มีความจำเป็นต้องออกไปทำกิจกรรมในพื้นที่ไกลๆ เพื่อลดต้นทุนในการบริหาร
2. การใช้บริการทางการตลาดออนไลน์ เช่น Google Ads, Facebook Ads, Instagram Ads เป็นต้น เพื่อให้สถาบันการศึกษาประชาสัมพันธ์ตัวเองและเสนอข้อมูลเกี่ยวกับสถาบันการศึกษาให้กับกลุ่มเป้าหมายอย่างได้ผล และไปยังพื้นที่ที่ไม่ได้เป็นกลุ่มลูกค้าในอดีต
3. ติดต่อกับโรงเรียนในพื้นที่ใกล้เคียงจังหวัดสงขลา โดยติดต่อกับโรงเรียนท้องถิ่นและโรงเรียนมัธยมปลาย เพื่อเสนอข้อมูลเกี่ยวกับสถาบันการศึกษา

องค์ความรู้ที่ได้จากการวิจัย (Research Knowledge)

การวิจัยนี้ได้เสนอเทคนิคการเหมืองข้อมูล (Data Mining) เพื่อเพิ่มประสิทธิภาพของการวางแผนนโยบายเชิงกลยุทธ์สถาบันการศึกษาระดับอุดมศึกษา โดยหากนำเทคนิคการเหมืองข้อมูล เช่นการจัดกลุ่ม (Clustering) มาใช้ในกระบวนการการศึกษาระดับสูง จะช่วยเพิ่มประสิทธิภาพในการดำเนินการรับสมัครกลุ่มนักศึกษาได้ตามเป้าหมาย จะช่วยลดค่าใช้จ่ายในการดำเนินกิจกรรมนี้ นอกจากนี้การวิเคราะห์ข้อมูลเป็นสิ่งสำคัญที่สำคัญสำหรับการสนับสนุนการตัดสินใจในทุกๆ อุตสาหกรรมไม่ว่าจะเป็นประเภทใด ส่วนสำคัญของงานวิจัยนี้คือการทบทวนบทบาทของเทคโนโลยีเหมืองข้อมูลในการวิเคราะห์ข้อมูล

ข้อเสนอแนะการวิจัย (Research Suggestions)

1. ข้อเสนอแนะการนำผลการวิจัยไปใช้

มหาวิทยาลัยสามารถนำไปใช้เรียนรู้ผลการเรียนของนักเรียน พฤติกรรมของนักเรียน ออกแบบหลักสูตรและการรับสมัครอย่างระมัดระวัง เพื่อส่งเสริมให้นักเรียนและนักศึกษาประสบความสำเร็จได้ง่ายขึ้น และสร้างกิจกรรมรับสมัครที่ตรงกลุ่มเป้าหมาย เพื่อลดงบประมาณสิ้นเปลืองที่ไม่ตรงกลุ่มเป้าหมาย

2. ข้อเสนอแนะการวิจัยครั้งต่อไป

2.1 เพิ่มขนาดของข้อมูล งานวิจัยในอนาคตจะทำการเพิ่มขนาดของข้อมูลที่ใช้ในการวิจัย เพื่อที่จะได้เห็นภาพรวมที่มีความแท้จริงของปัจจัยที่มีผลต่อการเลือกเข้าศึกษาในมหาวิทยาลัย โดยนำข้อมูลจากทุกคณะมาวิเคราะห์เพื่อให้เห็นภาพรวมมากขึ้น รวมทั้งการนำเข้าข้อมูลจากแหล่งที่มาต่างๆ เช่น ข้อมูลจากสื่อสังคมออนไลน์ ข้อมูลจากการสอบและทดสอบความสามารถ ข้อมูลเกี่ยวกับประสิทธิภาพของสถาบันการศึกษา เป็นต้น

2.2 วิเคราะห์ข้อมูลโดยใช้เทคนิคอื่นเพิ่มเติมและทำการเปรียบเทียบผลลัพธ์ เช่นนำเทคนิควิเคราะห์ข้อมูลอื่นๆ เช่น Decision tree หรือ Association Rule Mining เข้ามาใช้งานเพื่อเพิ่มความแม่นยำในการวิเคราะห์ข้อมูล

เอกสารอ้างอิง (References)

- สถิติการสมัครสอบ และชำระเงินค่าสมัครสอบ GAT/PAT. (2566). [ออนไลน์]. แหล่งที่มา: <https://www.dek-d.com/tcas/61505/> สืบค้นเมื่อ 20 เมษายน 2566.
- สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม (2566). **จำนวนนักศึกษาใหม่ ปีการศึกษา 2564 ภาคเรียนที่ 1 จำแนกตาม จังหวัด.** [ออนไลน์]. แหล่งที่มา: https://info.mhesi.go.th/stat_std_new.php?search_year=2564&download=7085&file_id=202208151109.xlsx. สืบค้นเมื่อ 25 เมษายน 2566.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*. 113: 177-194.
- Ayele, W. Y. (2020). Adapting CRISP-DM for idea mining: a data mining process for generating ideas using a textual dataset. *International Journal of Advanced Computer Sciences and Applications*. 11(6): 20-32.
- Celebi, M. E. (Ed.). (2014). *Partitional clustering algorithms*. Springer. 147-192

- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. (2): 224-227.
- Goyal, M., & Vohra, R. (2012). Applications of data mining in higher education. **International Journal of Computer Science Issues (IJCSI)**. 9(2): 113-120.
- King, R. S. (2015). **Cluster analysis and data mining: An introduction**. Mercury Learning and Information.
- Larose, D. T., & Larose, C. D. (2014). **Discovering knowledge in data: an introduction to data mining** (Vol. 4). John Wiley & Sons. 149.
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. **IEEE Transactions on Knowledge and Data Engineering**. 33(8): 3048-3061.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. **Journal of Educational Technology & Society**. 17(4): 49-64.

