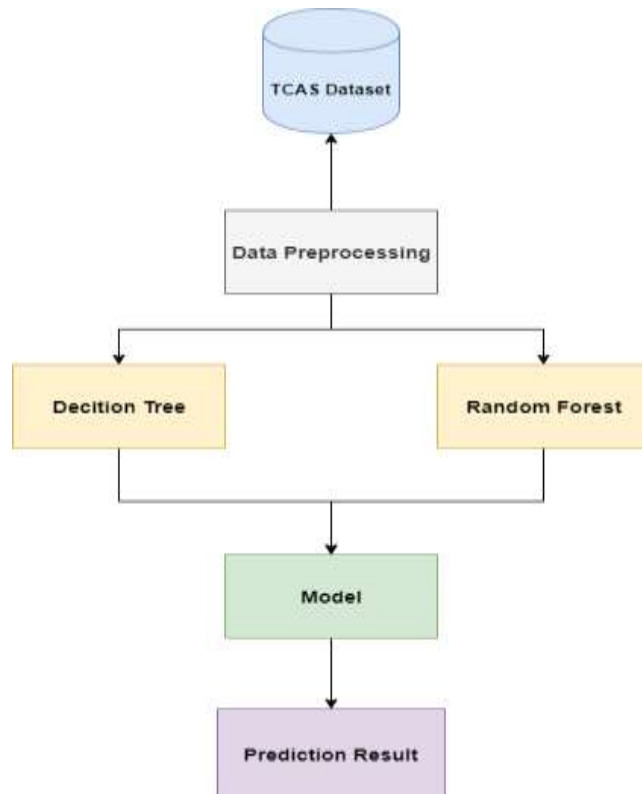


บทที่ 3

วิธีการดำเนินงานโครงการ

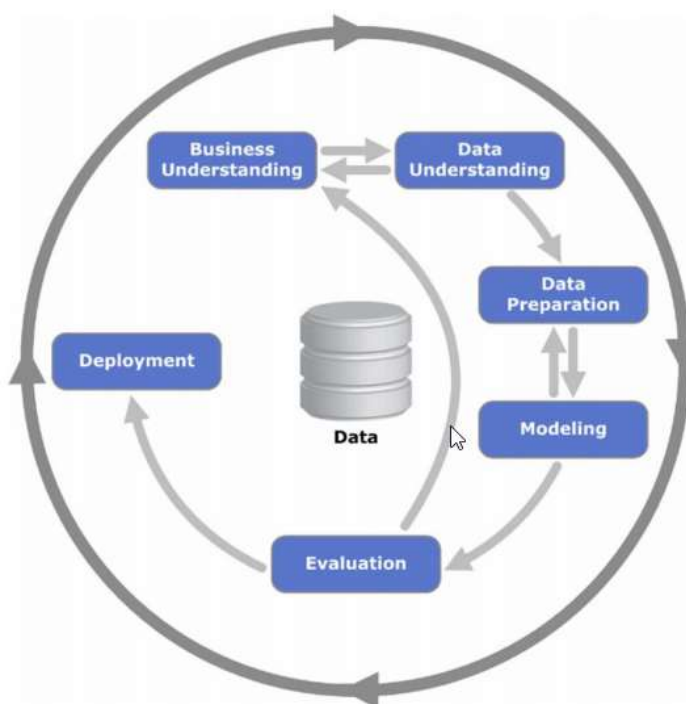
โครงการนี้ผู้จัดทำได้นำเสนอเทคนิคการทำเหมืองข้อมูล (Data mining) ด้วยวิธีการจำแนกข้อมูล (Data Classification) แบบวิธีต้นไม้ตัดสินใจ (Decision tree) เปรียบเทียบกับวิธีการจำแนกข้อมูลแบบ Random Forest เพื่อวิเคราะห์หาสารสนเทศที่เกิดขึ้นจากข้อมูลนักเรียนผู้มีสิทธิ์ Clearing house นำรูปแบบที่ให้ผลพยากรณ์ที่แม่นยำที่สุดได้มาวางแผนจำนวนผู้ผ่านการคัดเลือกเพื่อมีสิทธิ์ในการ ยืนยันสิทธิ์ Clearing house ในระบบ TCAS โดยข้อมูลที่ใช้ในการวิเคราะห์ครั้งนี้ คือข้อมูลผู้ผ่านการคัดเลือกที่มีสิทธิ์ ในการยืนยันสิทธิ์ รอบที่ 2-3 การรับแบบโควตาที่มีการสอบข้อเขียนหรือข้อสอบปฏิบัติ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ประจำปีการศึกษา 2563 จำนวน 3,195 คน โดยการแบ่งข้อมูลเป็น 2 กลุ่มคือ ข้อมูลทดสอบ (Data Testing) และข้อมูลเรียนรู้ (Data Training) โดยใช้หลักการแบ่งข้อมูลแบบ 10-Fold Cross Validation ผู้จัดทำได้แสดงขั้นตอน และกรอบแนวคิดการวิจัยเพื่อให้ทราบกระบวนการในการวิจัยในครั้งนี้โดยมีรายละเอียดดังรูปภาพที่ 3.1



ภาพที่ 3.1 กรอบแนวคิดขอบเขตโครงการ

ผู้จัดทำได้เลือกข้อมูลที่เกี่ยวข้องกับการทดลองซึ่งเป็นข้อมูลที่ใช้ในการประกอบการพิจารณาผู้สมัครเข้าศึกษาในมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ได้แก่ ข้อมูลขนาดสถานศึกษา, จังหวัดของสถานศึกษา, คะแนนมาตรฐานความถนัดทั่วไป (GAT), การทดสอบการศึกษาในระดับชาติขั้นพื้นฐานระดับ มัธยมศึกษาปีที่ 6 (ONET), เกรดเฉลี่ยสะสม 6 ภาคเรียน, คะแนนรวม และผลการยืนยันสิทธิ์ เพื่อวิเคราะห์หารูปแบบที่เกิดขึ้นจากการยืนยันสิทธิ์ และนำสารสนเทศที่ได้มาวางแผนการคัดเลือกบุคคล เข้าศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา นำไปสู่การวางแผนประชาสัมพันธ์ต่อไปในอนาคต

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM



ภาพที่ 3.2 กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM

ที่มา : sajeegm301.blogspot.com (ม.ป.ป.)

กระบวนการวิเคราะห์ข้อมูล (Cross Industry Standard Process for Data Mining หรือ CRISP. DM) พัฒนาขึ้นในปี ค.ศ. 1936 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ซึ่งกระบวนการวิเคราะห์ข้อมูล CRISP-UM. จะเป็นดังรูปแบบในรูปภาพที่ 3.2 แต่ละขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือ ขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกระบวนการนั้น ๆ ประกอบด้วย 6 ขั้นตอนหลักดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของการกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม คณะผู้จัดทำได้ทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง ดาต้า ไม่นิ่ง และวางแผนในการดำเนินการโดยการวิเคราะห์ข้อมูลรูปแบบความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลของข้อมูลนักศึกษา และข้อมูลการรับสมัครนักศึกษาโดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ ข้อมูลรับสมัครนักศึกษาเพื่อวางแผนการคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา นำไปสู่การวางแผนประชาสัมพันธ์ปีการศึกษา 2563 จำนวนรายการทั้งหมด 3,195 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูลจากสำนักส่งเสริมวิชาการและงานทะเบียน ตลอดจนการพิจารณาตรวจสอบดูความถูกต้องและความครบถ้วนของข้อมูล ที่ได้รับ โดยพิจารณาข้อมูลทั้งหมดว่าส่วนใดของข้อมูลที่จำเป็นต่อการวิเคราะห์ข้อมูลให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้คณะผู้จัดทำทำการรวบรวมข้อมูล เพื่อตรวจสอบรายละเอียด GPA และสถาบันการศึกษาเดิมของข้อมูลรับสมัครนักศึกษา ที่ได้จากฐานข้อมูลของมหาวิทยาลัยจากปีการศึกษา 2563 ซึ่งข้อมูลรับสมัครนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ปีการศึกษา 2563 มี 4,172 รายการ ประกอบด้วย 22 แอดมิททิวิตี ประกอบด้วย เลขที่ที่นั่งสอบ รหัสหลักสูตรที่สมัคร หลักสูตรที่สมัคร คณะที่สมัคร พื้นที่ที่สมัคร สถานศึกษาเดิม จังหวัด วุฒิเดิม เกรดเฉลี่ย ข้อมูลที่อยู่: อำเภอ ข้อมูลที่อยู่:จังหวัด รหัสหลักสูตรที่สอบผ่าน หลักสูตรที่สอบผ่าน คณะที่สอบผ่าน พื้นที่ที่สอบผ่าน ข้อมูลขนาดสถานศึกษา, จังหวัดของสถานศึกษา, คะแนนมาตรฐานความถนัดทั่วไป (GAT), การทดสอบการศึกษาระดับชาติขั้นพื้นฐานระดับ มัธยมศึกษาปีที่ 6 (ONET), เกรดเฉลี่ยสะสม 6 ภาคเรียน, คะแนนรวม และผลการยืนยันสิทธิ์

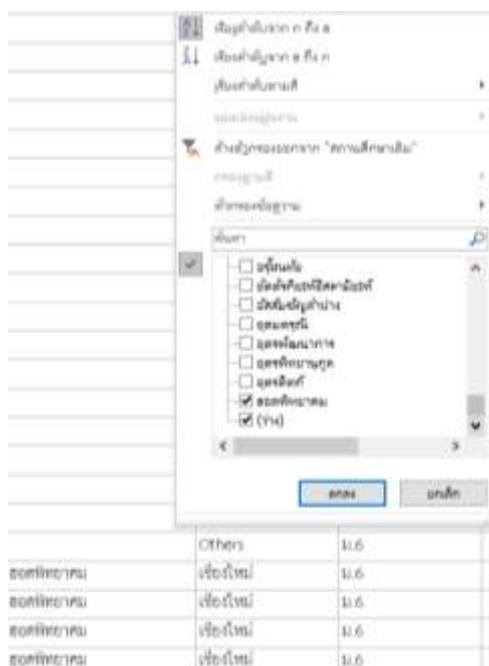
A	B	C	D	E	F	G
School_name	School_province	gpax6_score	onet_AVG	GAT	score	statustcas
กาฬสินธุ์พิทยาสรรพ์	Kalasin	2.39	32.49	43	16.25	ยื่นยื่นสิทธิ์
จักรคำคณาทร จังหวัดลำพูน	Lamphun	3.4	31.95	28.95	35.68	ยื่นยื่นสิทธิ์
เชียงใหม่วิทยาคม	Chiang Rai	3.72	39.58	32.49	37.06	ยื่นยื่นสิทธิ์
เทพศิรินทร์วิทยา	Chiang Mai	2.95	32.3	31.95	36.8	ยื่นยื่นสิทธิ์
ปายวิทยาคาร	Mae Hong Son	2.6	31.33	22	21.82	ยื่นยื่นสิทธิ์
แม่ออนวิทยาลัย (ฮอนเหนือ)	Chiang Mai	3.17	41.6	32.8	20.8	ยื่นยื่นสิทธิ์
อุพราษวิทยาลัย	Chiang Mai	3.47	30.5	23.7	31.38	ยื่นยื่นสิทธิ์
อุพราษวิทยาลัย	Chiang Mai	3.3	26.6	39.58	27.4	ยื่นยื่นสิทธิ์
อุพราษวิทยาลัย	Chiang Mai	3.27	28.6	42.7	22.45	ยื่นยื่นสิทธิ์
อุพราษวิทยาลัย	Chiang Mai	3.12	30.8	29.05	22.7	ยื่นยื่นสิทธิ์
ลำปางกัลยาณี	Lampang	2.95	34.8	29	39.1	ยื่นยื่นสิทธิ์
วัฒโนทัยพายัพ	Chiang Mai	2.86	25.9	31.5	26.38	ยื่นยื่นสิทธิ์
วัฒโนทัยพายัพ	Chiang Mai	3.34	30.8	32.3	30.33	ยื่นยื่นสิทธิ์
สวนบุญญูปถัมภ์ ลำพูน	Lamphun	2.6	26.15	31.25	32.38	ยื่นยื่นสิทธิ์
สวนบุญญูปถัมภ์ ลำพูน	Lamphun	3.54	28.7	35.1	36.5	ยื่นยื่นสิทธิ์
สารสาสน์วิเทศเชียงใหม่	Chiang Mai	2.89	31.6	27.85	32.65	ยื่นยื่นสิทธิ์
ปิ่นสร้อยแยดส่ววิทยาลัย	Chiang Mai	2.55	57.8	31.33	58.15	ยื่นยื่นสิทธิ์
วังชัยวิทยา	Chiang Mai	2.98	43.25	26.95	48.69	ยื่นยื่นสิทธิ์

ภาพที่ 3.4 ข้อมูลที่ผ่านการ Data Cleaning

3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางคณะผู้จัดทำได้ดำเนินการ ดังนี้

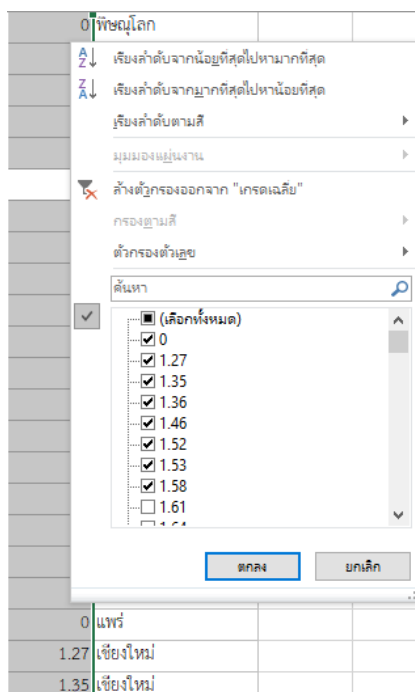
ข้อมูลรับสมัครนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ปีการศึกษา 2563 คณะผู้จัดทำได้ทำการลบข้อมูลที่เป็นช่องว่าง ซึ่งคณะผู้จัดทำพบว่า บางข้อมูลที่นักศึกษาไม่ได้กรอกหรือกรอกได้ไม่ถูกต้องตามที่ทางระบบได้กำหนดไว้ ดังนั้นคณะผู้จัดทำได้ทำการลบข้อมูลที่สำคัญต่อการนำมาวิเคราะห์ ดังนี้

คณะผู้จัดทำได้ทำการตรวจสอบข้อมูลที่เป็นข้อมูลช่องว่างและข้อมูลที่ผิดพลาดพบข้อมูลที่ช่องว่างและข้อมูลที่ผิดพลาด จำนวน 360 รายการ และคัดข้อมูลสถานะผู้สมัคร “ไม่ผ่านการคัดเลือก” ออกอีกจำนวน 617 รายการ ดังนั้นจะมีข้อมูลรับสมัครนักศึกษา ประจำปี 2563 ที่ทำการ Cleaning แล้ว จำนวน 3,195 รายการ สำหรับทำการวิเคราะห์



ภาพที่ 3.5 การตรวจสอบข้อมูลที่เป็นค่าว่างและข้อมูลผิดพลาดทั้งหมด

แอดทริบิวต์ เกรตเฉลี่ย มีค่าเป็น 0 จึงได้ทำการตรวจสอบข้อมูลและทำการลบแถวนั้นทิ้ง



ภาพที่ 3.6 การตรวจสอบข้อมูลที่เป็นค่าเป็น 0 ทั้งหมด

3.1.3.3 การแปลงรูปแบบของข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูล โดยการจัดกลุ่มข้อมูลที่เป็นตัวเลขให้อยู่ในกลุ่มต่าง ๆ ตามหลักทางสถิติและกลุ่มตัวอย่าง จากข้อมูลที่ใช้ในการทดลองผู้จัดทำได้ทำการแทนค่าให้กับข้อมูลก่อนการวิเคราะห์ทดลองดังต่อไปนี้

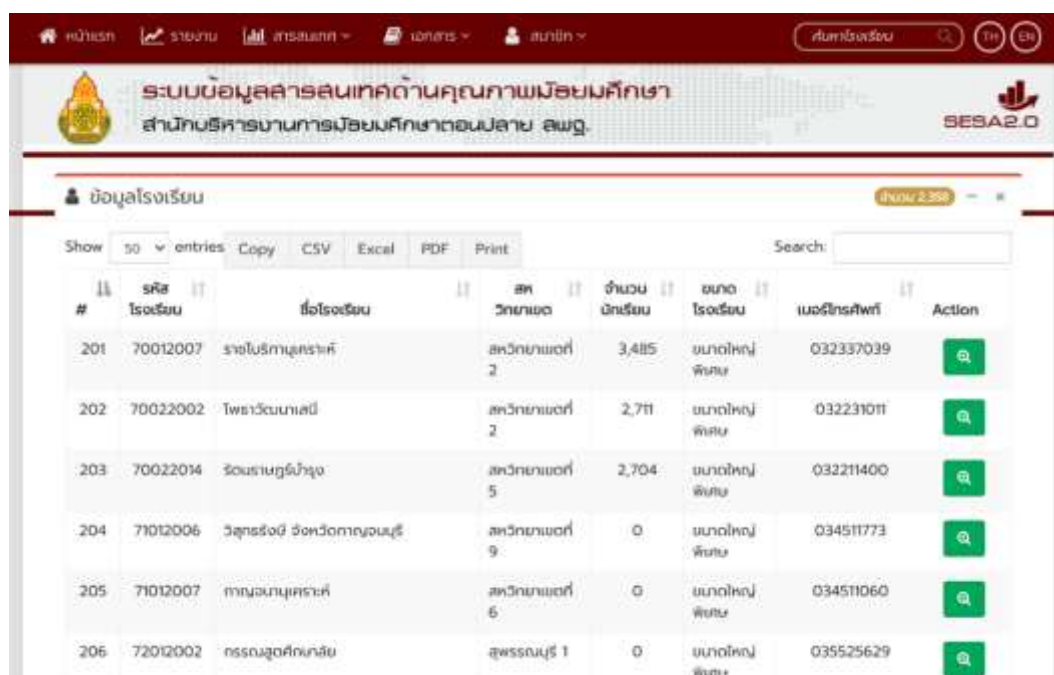
1) ข้อมูลขนาดโรงเรียนผู้จัดทำได้ใช้เกณฑ์การแบ่งขนาดโดยยึดตามข้อมูลสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) ดังนี้

โรงเรียนขนาดเล็กมีนักเรียน 1 – 499 คน แทนค่า (S),

โรงเรียนขนาดกลาง มีนักเรียน 500 – 1,499 คน แทนค่า (M),

โรงเรียนขนาดใหญ่ มีนักเรียน 1,500 – 2,400 คน แทนค่า (L),

โรงเรียนขนาดใหญ่พิเศษมีนักเรียน 2,500 คนขึ้นไป แทนค่า (XL)



The screenshot shows the 'ระบบข้อมูลสารสนเทศด้านคุณภาพมาตรฐานการศึกษา' (Education Quality Standard Information System) interface. It displays a table with columns for school ID, school name, province, number of students, school type, and phone number. The table contains 6 rows of data.

#	รหัสโรงเรียน	ชื่อโรงเรียน	ศท วิทยาเขต	จำนวน นักเรียน	ขนาด โรงเรียน	เบอร์โทรศัพท์	Action
201	70012007	โรงเรียนพุทธมณฑลราชวิถี	สทวิทยาเขต 2	3,485	ขนาดใหญ่พิเศษ	032337039	[Icon]
202	70022002	โรงเรียนสวนกุหลาบวิทยาลัย	สทวิทยาเขต 2	2,711	ขนาดใหญ่พิเศษ	032231011	[Icon]
203	70022014	โรงเรียนสวนกุหลาบวิทยาลัย	สทวิทยาเขต 5	2,704	ขนาดใหญ่พิเศษ	032211400	[Icon]
204	71012006	วิทยาลัยสงฆ์ วัดหนองตาครุ	สทวิทยาเขต 9	0	ขนาดใหญ่พิเศษ	034511773	[Icon]
205	71012007	วิทยาลัยสงฆ์ วัดหนองตาครุ	สทวิทยาเขต 6	0	ขนาดใหญ่พิเศษ	034511060	[Icon]
206	72012002	วิทยาลัยสงฆ์ วัดหนองตาครุ	สทวิทยาเขต 1	0	ขนาดใหญ่พิเศษ	035525629	[Icon]

ภาพที่ 3.7 เว็บไซต์ข้อมูลสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) ดังนี้

2) ข้อมูลด้านคะแนนมาตรฐานความรู้ทั่วไป (GAT), การทดสอบการศึกษา ระดับชาติขั้นพื้นฐานระดับ มัธยมศึกษาปีที่ 6 (ONET), และข้อมูลคะแนนรวม นั้นเป็นข้อมูลแบบทศนิยมต่อเนื่องข้อมูลจะมีการกระจาย เพื่อลดการกระจายของข้อมูลผู้จัดทำได้แบ่งช่วงของข้อมูล (Binning data) คณะผู้จัดทำได้แบ่งช่วงของข้อมูลคะแนนออกเป็น 3 ช่วงคะแนน คือ

ช่วงคะแนนระหว่าง 0.0000 – 35.0000 อยู่ในระดับต่ำ แทนค่า (Low),
ช่วงคะแนนระหว่าง 35.0001 – 55.0000 อยู่ในระดับปานกลาง แทนค่า
(Moderate),

ช่วงคะแนนระหว่าง 55.0001 – 100 อยู่ในระดับสูง แทนค่า (High)

3) ข้อมูลจังหวัดของสถานศึกษา พิจารณาจากจังหวัดที่ตั้งของสถานศึกษา

ถ้าที่ตั้งอยู่ในจังหวัดเชียงใหม่ให้แทนค่า (inCM)

ถ้าที่ตั้งอยู่นอกจังหวัดเชียงใหม่ให้แทนค่า (outCM)

4) ข้อมูลผลการยืนยันสิทธิ์

ถ้าผู้สมัครยืนยันสิทธิ์ ให้แทนค่า (Confirm)

ถ้าผู้สมัครไม่ยืนยันสิทธิ์ให้แทนค่า (NotConfirm)

A	B	C	D	E	F	G
school_size	school_province	gpax6_score	onet_AVG	gatTotal	score	statuscas
M	inCM	High	Moderate	Moderate	Moderate	NotConfirm
M	inCM	High	Low	Moderate	Low	NotConfirm
XL	outCM	Moderate	Moderate	Low	Low	Confirm
XL	outCM	High	Moderate	Moderate	Moderate	Confirm
XL	outCM	Moderate	Low	Low	Low	NotConfirm
XL	outCM	Moderate	Moderate	Moderate	Low	NotConfirm
XL	outCM	High	Low	Low	Low	NotConfirm
M	outCM	High	Moderate	Low	Moderate	Confirm
XL	inCM	Moderate	Moderate	Moderate	Moderate	NotConfirm
XL	outCM	High	Low	Moderate	Low	NotConfirm
L	outCM	High	Low	Moderate	Low	NotConfirm
L	inCM	Moderate	Moderate	Moderate	Moderate	NotConfirm
L	inCM	Moderate	Moderate	Moderate	Moderate	Confirm
L	inCM	Low	Moderate	Moderate	Low	NotConfirm
L	outCM	Moderate	Moderate	Moderate	Moderate	NotConfirm
M	outCM	Moderate	Low	Moderate	Low	NotConfirm
M	outCM	Moderate	Moderate	Low	Low	Confirm
L	inCM	Moderate	Low	Low	Low	NotConfirm
XL	inCM	Moderate	Moderate	Moderate	Moderate	NotConfirm

ภาพที่ 3.8 เปลี่ยนข้อมูลเป็นภาษาอังกฤษ

3.1.4 สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และ สถิติ เพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ อาทิ การจำแนก (Classification) ต้นไม้ตัดสินใจ (Decision Tree) และ Random Forest

คณะผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Decision Tree และ Random Forest เพื่อ ใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่ได้มาวางแผนการคัดเลือกบุคคล เข้าศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา นำไปสู่การวางแผน ประชาสัมพันธ์ต่อไปในอนาคต

school_size	school_province	gpax6_score	ONET	GAT	Score	statustcas
M	inCM	High	Moderate	Moderate	Moderate	NoConfirm
M	inCM	High	Low	Low	Low	NoConfirm
XL	outCM	Medium	Low	Low	Low	Confirm
XL	outCM	High	Low	Moderate	Moderate	Confirm
XL	outCM	Medium	Low	Low	Low	NoConfirm
XL	outCM	Medium	Low	Low	Low	NoConfirm
XL	outCM	High	Low	Low	Low	NoConfirm
M	outCM	High	Moderate	Low	Moderate	Confirm
XL	inCM	Medium	Moderate	Moderate	Moderate	NoConfirm
XL	outCM	High	Low	Low	Low	NoConfirm
L	outCM	High	Low	Low	Low	NoConfirm
L	inCM	Medium	Low	Moderate	Moderate	NoConfirm
L	inCM	Medium	Low	Moderate	Moderate	Confirm
L	inCM	Low	Low	Low	Low	NoConfirm
L	outCM	Medium	Moderate	Moderate	Moderate	NoConfirm
M	outCM	Medium	Low	Low	Low	NoConfirm
M	outCM	Medium	Low	Low	Low	Confirm
L	inCM	Medium	Low	Low	Low	NoConfirm
XL	inCM	Medium	Moderate	Moderate	Moderate	NoConfirm
XL	inCM	Medium	Moderate	High	Moderate	NoConfirm
M	inCM	High	Moderate	Moderate	Moderate	NoConfirm
XL	outCM	High	Low	Moderate	Moderate	NoConfirm
XL	outCM	High	Moderate	Low	Low	NoConfirm
XL	outCM	Low	Low	Low	Low	NoConfirm

ภาพที่ 3.9 ชุดข้อมูลที่คัดเลือกมาวิเคราะห์ข้อมูล

จากรูปภาพที่ 3.9 ประกอบด้วย 7 แอตทริบิวต์ คือ

school_size (ขนาดของโรงเรียน) ประกอบด้วย 4 ค่า คือ S M L XL

school_province (ที่ตั้งของโรงเรียน) ประกอบด้วย 2 ค่า คือ inCM outCM

gpax6_score (เกรดเฉลี่ย 6 เทอม) ประกอบด้วย 3 ค่า คือ Low Medium High

ONET (คะแนน ONET) ประกอบด้วย 3 ค่า คือ Low Moderate High

GAT (คะแนน GAT) ประกอบด้วย 3 ค่า คือ Low Moderate High

Score ประกอบด้วย 3 ค่า คือ Low Moderate High

Statustcas ประกอบด้วย 3 ค่า คือ Confirm NotConfrim

3.1.5. การสร้างโมเดล Decision tree ด้วยการคำนวณ

การสร้างโมเดล Decision tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อยๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จากสมการดังนี้

ตารางที่ 3.1 สมการการหาความสัมพันธ์ของแอตทริบิวต์

การคำนวณ	สมการ
Entropy	$entropy(c1) = -p(c1) \log p(c1)$ และ $p(c2)$ คือ ค่าความน่าจะเป็นของ $c1$
Information Gain	$IG (parent, child) = entropy(parent) - [p(c1) \times entropy(c1) + p(c2) * entropy(C2) + \dots]$

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น Root ของ Decision Tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์ Confirm และ NotConfrim ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ school_size จากข้อมูลสามารถคำนวณค่า IG ได้
ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4896 \times \log_2 (0.4896) + 0.5104 \times \log_2 (0.5104)] \\ &= 0.9997 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=L)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4587 \times \log_2 (0.4587) + 0.5413 \times \log_2 (0.5413)] \\ &= 0.99507 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=M)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5016 \times \log_2 (0.5016) + 0.4984 \times \log_2 (0.4984)] \\ &= 0.99999 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=S)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.6026 \times \log_2 (0.6026) + 0.3974 \times \log_2 (0.3974)] \\ &= 0.96943 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=XL)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4874 \times \log_2 (0.4874) + 0.5104 \times \log_2 (0.5104)] \\ &= 0.99954 \end{aligned}$$

$$\begin{aligned}
 IG(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(\text{ผล} = L) \times (\text{ผล} = L) + p(\text{ผล} = M) \times (\text{ผล} = M) \\
 &+ p(\text{ผล} = S) \times (\text{ผล} = S) + p(\text{ผล} = XL) \times (\text{ผล} = XL)] \\
 &= 0.9997 - [0.2223 \times 0.99507 + 0.2020 \times 0.99999 + \\
 &0.0494 \times 0.96943 + 0.5263 \times 0.99954] \\
 &= 0.9997 - [0.2212 + 0.2020 + 0.479 + 0.5260] \\
 &= 0.9997 - 0.9972 \\
 &= 0.00253
 \end{aligned}$$

2) คำนวณค่า IG ของแอตทริบิวต์ school_province จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
 \text{entropy}(\text{parent}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4896 \times \log_2 (0.4896) + 0.5104 \times \log_2 (0.5104)] \\
 &= 0.9997
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{ผล} = \text{inCM}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5197 \times \log_2 (0.5197) + 0.4803 \times \log_2 (0.4803)] \\
 &= 0.99888
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{ผล} = \text{outCM}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4576 \times \log_2 (0.4576) + 0.5424 \times \log_2 (0.5424)] \\
 &= 0.99481
 \end{aligned}$$

$$IG(\text{parent, child}) = \text{entropy}(\text{parent}) - [p(\text{ผล} = \text{inCM}) \times (\text{ผล} = \text{inCM}) + p(\text{ผล} = \text{outCM}) \times (\text{ผล} = \text{outCM})]$$

$$\begin{aligned}
&= 0.9997 - [0.5142 \times 0.99888 + 0.4858 \times 0.99481] \\
&= 0.9997 - [0.5137 \times 0.4832] \\
&= 0.9997 - 0.9969 \\
&= 0.00278
\end{aligned}$$

3) คำนวณค่า IG ของแอตทริบิวต์ gpa6_score จากข้อมูลสามารถคำนวณค่า IG ได้
ดังนี้

$$\begin{aligned}
\text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
&\log_2 p(\text{NotConfirm}) \\
&= -[0.4896 \times \log_2 (0.4896) + 0.5104 \times \log_2 (0.5104)] \\
&= 0.9997
\end{aligned}$$

$$\begin{aligned}
\text{entropy (ผล=High)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
&\log_2 p(\text{NotConfirm}) \\
&= -[0.4350 \times \log_2 (0.4350) + 0.5650 \times \log_2 (0.5650)] \\
&= 0.98776
\end{aligned}$$

$$\begin{aligned}
\text{entropy (ผล=Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
&\log_2 p(\text{NotConfirm}) \\
&= -[0.5827 \times \log_2 (0.5827) + 0.4173 \times \log_2 (0.4173)] \\
&= 0.98019
\end{aligned}$$

$$\begin{aligned}
\text{entropy (ผล=Medium)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
&\log_2 p(\text{NotConfirm}) \\
&= -[0.5237 \times \log_2 (0.5237) + 0.4763 \times \log_2 (0.4763)] \\
&= 0.99838
\end{aligned}$$

$$\begin{aligned}
 IG(\text{parent, child}) &= \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Medium}) \times (\text{ผล=Medium})] \\
 &= 0.9997 - [0.4383 \times 0.98776 + 0.0804 \times 0.98019 + 0.4813 \times 0.99838] \\
 &= 0.9997 - [0.4329 \times 0.0788 \times 0.4805] \\
 &= 0.9997 - 0.9923 \\
 &= 0.00742
 \end{aligned}$$

4) คำนวณค่า IG ของแอดทริบิวต์ ONET จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\begin{aligned}
 \text{entropy}(\text{parent}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm}) \\
 &= -[0.4896 \times \log_2(0.4896) + 0.5104 \times \log_2(0.5104)] \\
 &= 0.9997
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{ผล=High}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm}) \\
 &= -[0.1818 \times \log_2(0.1818) + 0.8182 \times \log_2(0.8182)] \\
 &= 0.68404
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{ผล=Low}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm}) \\
 &= -[0.5302 \times \log_2(0.5302) + 0.4698 \times \log_2(0.4698)] \\
 &= 0.99737
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy}(\text{ผล=Moderate}) &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm}) \\
 &= -[0.4505 \times \log_2(0.4505) + 0.5495 \times \log_2(0.5495)]
 \end{aligned}$$

$$= 0.99291$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.9997 - [0.0070 \times 0.68404 + 0.5136 \times 0.99737 + 0.4794 \times 0.99291]$$

$$= 0.9997 - [0.0048 \times 0.5123 \times 0.4760]$$

$$= 0.9997 - 0.9930$$

$$= 0.00664$$

5) คำนวณค่า IG ของแอตทริบิวต์ GAT จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4896 \times \log_2 (0.4896) + 0.5104 \times \log_2 (0.5104)]$$

$$= 0.9997$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.2286 \times \log_2 (0.2286) + 0.7714 \times \log_2 (0.7714)]$$

$$= 0.77551$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5201 \times \log_2 (0.5201) + 0.4799 \times \log_2 (0.4799)]$$

$$= 0.99883$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4685 \times \log_2 (0.4685) + 0.5315 \times \log_2 (0.5315)]$$

$$= 0.99714$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.9997 - [0.0443 \times 0.77551 + 0.6137 \times 0.99883 + 0.3420 \times 0.99714]$$

$$= 0.9997 - [0.03438 \times 0.61296 \times 0.34101]$$

$$= 0.9997 - 0.988$$

$$= 0.01133$$

6) คำนวณค่า IG ของแอตทริบิวต์ score จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4896 \times \log_2 (0.4896) + 0.5104 \times \log_2 (0.5104)]$$

$$= 0.9997$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.1538 \times \log_2 (0.1538) + 0.8462 \times \log_2 (0.8462)]$$

$$= 0.61938$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5432 \times \log_2 (0.5432) + 0.4568 \times \log_2 (0.4568)]$$

$$= 0.99460$$

$$\begin{aligned} \text{entropy (ผล=Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4165 \times \log_2 (0.4165) + 0.5835 \times \log_2 (0.5835)] \\ &= 0.97980 \end{aligned}$$

$$\begin{aligned} \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=} \\ &\text{Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})] \\ &= 0.9997 - [0.0082 \times 0.61938 + 0.5934 \times 0.99460 + \\ &0.3984 \times 0.97980] \\ &= 0.9997 - [0.0051 \times 0.59021 \times 0.39031] \\ &= 0.9997 - 0.986 \\ &= 0.01407 \end{aligned}$$

จากการคำนวณค่า IG ของทุกแอดทริบิวต์พบว่าค่า IG ของแอดทริบิวต์ score มีค่ามากที่สุด (0.1407) ดังนั้นจึงเลือกแอดทริบิวต์ score ขึ้นมาเป็นโหนด root และจะต้องทำการแตกกิ่งจากโหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน และคณะผู้จัดทำพบว่า การคำนวณแอดทริบิวต์ High ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอดทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์เป็น Confirm และ NotConfirm ดังนั้นคณะผู้จัดทำจึงทำการสร้างโหนดในระดับถัดไปของแอดทริบิวต์ score

การคำนวณค่าแต่ละแอดทริบิวต์ในระดับที่ 2 ต่อจากโหนด root เพื่อหาค่า IG ที่มากที่สุดของแอดทริบิวต์ score กับแอดทริบิวต์ schoo_size school_province gpa6_score ONET และGAT จำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น Confirm และ NotConfirm ดังนี้

1) คำนวณค่า IG ของแอดทริบิวต์ score “Moderate” ต่อ school_province ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4165 \times \log_2 (0.4165) + 0.5835 \times \log_2 (0.5835)] \end{aligned}$$

$$= 0.9798$$

$$\text{entropy (ผล = inCM)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4194 \times \log_2 (0.4194) + 0.5806 \times \log_2 (0.5806)]$$

$$= 0.98119$$

$$\text{entropy (ผล = outCM)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4126 \times \log_2 (0.4126) + 0.5874 \times \log_2 (0.5874)]$$

$$= 0.97787$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=inCM}) \times (\text{ผล=inCM}) + p(\text{ผล=outCM}) \times (\text{ผล=outCM})]$$

$$= 0.9798 - [0.5723 \times 0.98119 + 0.4277 \times 0.97787]$$

$$= 0.9798 - [0.5616 \times 0.4182]$$

$$= 0.9798 - 0.9798$$

$$= 0.00003$$

2) คำนวณค่า IG ของแอตทริบิวต์ score “Moderate” ต่อ gpax6_score ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4165 \times \log_2 (0.4165) + 0.5835 \times \log_2 (0.5835)]$$

$$= 0.9798$$

$$\text{entropy (ผล = High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.3852 \times \log_2 (0.3852) + 0.6148 \times \log_2 (0.6148)]$$

$$= 0.96166$$

$$\text{entropy (ผล = Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.60 \times \log_2 (0.60) + 0.40 \times \log_2 (0.40)]$$

$$= 0.97095$$

$$\text{entropy (ผล = Medium)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4485 \times \log_2 (0.4485) + 0.5514 \times \log_2 (0.5514)]$$

$$= 0.99235$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Medium}) \times (\text{ผล=Medium})]$$

$$= 0.9798 - [0.5819 \times 0.96166 + 0.0318 \times 0.97095 + 0.3863 \times 0.99235]$$

$$= 0.9798 - [0.5596 \times 0.0309 \times 0.3834]$$

$$= 0.9798 - 0.9738$$

$$= 0.00599$$

3) คำนวณค่า IG ของแอดทริบิวต์ score “Moderate” ต่อ ONET ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4165 \times \log_2 (0.4165) + 0.5835 \times \log_2 (0.5835)]$$

$$= 0.9798$$

$$\begin{aligned}
 \text{entropy (ผล = High)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.1429 \times \log_2 (0.1429) + 0.8517 \times \log_2 (0.8517)] \\
 &= 0.59167
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล = Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4037 \times \log_2 (0.4037) + 0.5963 \times \log_2 (0.5963)] \\
 &= 0.97306
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล = Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4230 \times \log_2 (0.4230) + 0.5770 \times \log_2 (0.5770)] \\
 &= 0.98283
 \end{aligned}$$

$$\begin{aligned}
 \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=} \\
 &\text{Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})] \\
 &= 0.9798 - [0.0111 \times 0.59167 + 0.1733 \times 0.97306 + \\
 &0.8156 \times 0.98283] \\
 &= 0.9798 - [0.0066 \times 0.1686 \times 0.8016] \\
 &= 0.9798 - 0.9768 \\
 &= 0.00303
 \end{aligned}$$

4) คำนวณค่า IG ของแอดทริบิวต์ score “Moderate” ต่อ GAT ได้ดังนี้

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4165 \times \log_2 (0.4165) + 0.5835 \times \log_2 (0.5835)]
 \end{aligned}$$

$$= 0.9798$$

$$\text{entropy (ผล = High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.2586 \times \log_2 (0.2586) + 0.7414 \times \log_2 (0.7414)]$$

$$= 0.82466$$

$$\text{entropy (ผล = Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4028 \times \log_2 (0.4028) + 0.5972 \times \log_2 (0.5972)]$$

$$= 0.97255$$

$$\text{entropy (ผล = Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4426 \times \log_2 (0.4426) + 0.5574 \times \log_2 (0.5574)]$$

$$= 0.99048$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})]$$

$$= 0.9798 - [0.0922 \times 0.82466 + 0.2289 \times 0.97255 + 0.6789 \times 0.99048]$$

$$= 0.9798 - [0.0760 \times 0.2227 \times 0.6724]$$

$$= 0.9798 - 0.9711$$

$$= 0.00872$$

5) คำนวณค่า IG ของแอตทริบิวต์ score “Low” ต่อ school_size ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5432 \times \log_2 (0.5432) + 0.4568 \times \log_2 (0.4568)]$$

$$= 0.9946$$

$$\text{entropy (ผล=L)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4764 \times \log_2 (0.4764) + 0.5236 \times \log_2 (0.5236)]$$

$$= 0.99840$$

$$\text{entropy (ผล=M)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5190 \times \log_2 (0.5190) + 0.4810 \times \log_2 (0.4810)]$$

$$= 0.99895$$

$$\text{entropy (ผล=S)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6721 \times \log_2 (0.6721) + 0.3279 \times \log_2 (0.3279)]$$

$$= 0.91273$$

$$\text{entropy (ผล=XL)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5642 \times \log_2 (0.5642) + 0.4358 \times \log_2 (0.4358)]$$

$$= 0.98807$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=L}) \times (\text{ผล=L}) + p(\text{ผล=M}) \times (\text{ผล=M}) + p(\text{ผล=S}) \times (\text{ผล=S}) + p(\text{ผล=XL}) \times (\text{ผล=XL})]$$

$$= 0.9946 - [0.2038 \times 0.99840 + 0.2241 \times 0.99895 + 0.0651 \times 0.91273 + 0.5069 \times 0.98807]$$

$$= 0.9946 - [0.2035 + 0.2239 + 0.0594 + 0.5009]$$

$$= 0.9946 - 0.9877$$

$$= 0.00689$$

6) คำนวณค่า IG ของแอตทริบิวต์ score “Low” ต่อ school_province ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5432 \times \log_2 (0.5432) + 0.4568 \times \log_2 (0.4568)]$$

$$= 0.9946$$

$$\text{entropy (ผล=inCM)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6059 \times \log_2 (0.6059) + 0.3941 \times \log_2 (0.3941)]$$

$$= 0.96742$$

$$\text{entropy (ผล=outCM)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4868 \times \log_2 (0.4868) + 0.5132 \times \log_2 (0.5132)]$$

$$= 0.99950$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=inCM}) \times (\text{ผล=inCM}) + p(\text{ผล=outCM}) \times (\text{ผล=outCM})]$$

$$= 0.9946 - [0.4739 \times 0.96742 + 0.4858 \times 0.99950]$$

$$= 0.9946 - [0.4584 \times 0.4855]$$

$$= 0.9946 - 0.9439$$

$$= 0.05068$$

7) คำนวณค่า IG ของแอตทริบิวต์ score “Low” ต่อ gpax6_score ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5432 \times \log_2 (0.5432) + 0.4568 \times \log_2 (0.4568)] \\ &= 0.9946 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=High)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5016 \times \log_2 (0.5016) + 0.4984 \times \log_2 (0.4984)] \\ &= 0.99999 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5794 \times \log_2 (0.5794) + 0.4206 \times \log_2 (0.4206)] \\ &= 0.98171 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=Medium)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5616 \times \log_2 (0.5616) + 0.4384 \times \log_2 (0.4384)] \\ &= 0.98901 \end{aligned}$$

$$\begin{aligned} \text{IG (parent, child)} &= \text{entropy(parent)} - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=} \\ &\text{Low}) \times (\text{ผล= Low}) + p(\text{ผล= Medium}) \times (\text{ผล= Medium})] \\ &= 0.9946 - [0.3404 \times 0.99999 + 0.1142 \times 0.98171 + \\ &0.5454 \times 0.98901] \\ &= 0.9946 - [0.3404 \times 0.1121 \times 0.5394] \\ &= 0.9946 - 0.9919 \end{aligned}$$

$$= 0.00269$$

8) คำนวณค่า IG ของแอดทริบิวต์ score “Low” ต่อ ONET ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5432 \times \log_2 (0.5432) + 0.4568 \times \log_2 (0.4568)]$$

$$= 0.9946$$

$$\text{entropy (ผล = Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5499 \times \log_2 (0.5499) + 0.4501 \times \log_2 (0.4501)]$$

$$= 0.99282$$

$$\text{entropy (ผล = Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5234 \times \log_2 (0.5234) + 0.4766 \times \log_2 (0.4766)]$$

$$= 0.99842$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล} = \text{Low}) \times (\text{ผล} = \text{Low}) + p(\text{ผล} = \text{Moderate}) \times (\text{ผล} = \text{Moderate})]$$

$$= 0.9798 - [0.7492 \times 0.99282 + 0.2508 \times 0.99842]$$

$$= 0.9798 - [0.7438 \times 0.2504]$$

$$= 0.9798 - 0.9942$$

$$= 0.00038$$

9) คำนวณค่า IG ของแอดทริบิวต์ score “Low” ต่อ GAT ได้ดังนี้

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5432 \times \log_2 (0.5432) + 0.4568 \times \log_2 (0.4568)] \\
 &= 0.9946
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล = High)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0 \times \log_2 (0) + 1 \times \log_2 (1)] \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล = Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5406 \times \log_2 (0.5406) + 0.4594 \times \log_2 (0.4594)] \\
 &= 0.99524
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล = Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \\
 &\times \log_2 p(\text{NotConfirm}) \\
 &= -[0.5676 \times \log_2 (0.5676) + 0.4324 \times \log_2 (0.4324)] \\
 &= 0.98679
 \end{aligned}$$

$$\begin{aligned}
 \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) \\
 &+ p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})] \\
 &= 0.9798 - [0.0011 \times 0 + 0.8805 \times 0.99524 + 0.1185 \times \\
 &0.98679] \\
 &= 0.9798 - [0 \times 0.8763 \times 0.1169] \\
 &= 0.9798 - 0.9932 \\
 &= 0.00143
 \end{aligned}$$

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ school_province มีค่ามากที่สุด (0.5068) ดังนั้นจึงเลือกแอตทริบิวต์ school_province ขึ้นมาเป็นโน้ตระดับ 2 และจะต้องทำการแตกกิ่งจากโน้ตระดับ 3 ออกไปจนข้อมูลในแต่ละโน้ตมีคลาสค่าตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 ต่อจากระดับที่ 2 เพื่อหาค่า IG ที่มากที่สุดของแอตทริบิวต์ school_province ต่อ schoo_size gpa6_score ONET และ GAT กับจำนวนข้อมูล ทั้งหมดโดยใช้ผลลัพธ์เป็น Confirm และ NotConfirm ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ school_province “inCM” ต่อ schoo_size ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5197 \times \log_2 (0.5197) + 0.4803 \times \log_2 (0.4803)] \\ &= 0.99888 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=L)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5142 \times \log_2 (0.5142) + 0.4858 \times \log_2 (0.4858)] \\ &= 0.99942 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=M)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.5172 \times \log_2 (0.5172) + 0.4828 \times \log_2 (0.4828)] \\ &= 0.99914 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=S)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.8182 \times \log_2 (0.8182) + 0.1818 \times \log_2 (0.1818)] \end{aligned}$$

$$= 0.68404$$

$$\text{entropy (ผล=XL)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5159 \times \log_2 (0.5159) + 0.4841 \times \log_2 (0.4841)]$$

$$= 0.99927$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล= L}) \times (\text{ผล=L}) + p(\text{ผล= M}) \times (\text{ผล= M}) + p(\text{ผล= S}) \times (\text{ผล= S}) + p(\text{ผล= XL}) \times (\text{ผล= XL})]$$

$$= 0.99888 - [0.281532 \times 0.99942 + 0.216216 \times 0.99914 + 0.018018 \times 0.68404 + 0.484234 \times 0.99927]$$

$$= 0.99888 - [0.281368 + 0.216031 + 0.012325 + 0.483881]$$

$$= 0.99888 - 0.9936$$

$$= 0.005274$$

2) คำนวณค่า IG ของแอตทริบิวต์ school_province “inCM” ต่อ gpax6_score ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5197 \times \log_2 (0.5197) + 0.4803 \times \log_2 (0.4803)]$$

$$= 0.99888$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4455 \times \log_2 (0.4455) + 0.5545 \times \log_2 (0.5545)]$$

$$= 0.9914$$

$$\begin{aligned}
 \text{entropy (ผล=Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.7237 \times \log_2 (0.7237) + 0.2763 \times \log_2 (0.2763)] \\
 &= 0.85038
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5419 \times \log_2 (0.5419) + 0.4581 \times \log_2 (0.4581)] \\
 &= 0.99494
 \end{aligned}$$

$$\begin{aligned}
 \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=} \\
 &\text{Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})] \\
 &= 0.99888 - [0.290541 \times 0.9914 + 0.141892 \times \\
 &0.85038 + 0.567568 \times 0.99494] \\
 &= 0.99888 - [0.288041 \times 0.120662 \times 0.564693] \\
 &= 0.9946 - 0.9734 \\
 &= 0.025483
 \end{aligned}$$

3) คำนวณค่า IG ของแอตทริบิวต์ school_province “inCM” ต่อ ONET ได้ดังนี้

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5197 \times \log_2 (0.5197) + 0.4803 \times \log_2 (0.4803)] \\
 &= 0.99888
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=High)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.2857 \times \log_2 (0.2857) + 0.7143 \times \log_2 (0.7143)]
 \end{aligned}$$

$$= 0.86312$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6015 \times \log_2 (0.6015) + 0.3985 \times \log_2 (0.3985)]$$

$$= 0.97006$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4824 \times \log_2 (0.4824) + 0.5176 \times \log_2 (0.5176)]$$

$$= 0.9991$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.99888 - [0 \times 0.86312 + 0.97006 \times 0.85038 + 0.658409 \times 0.9991]$$

$$= 0.99888 - [0 \times 0.326306 \times 0.657819]$$

$$= 0.9946 - 0.9841$$

$$= 0.014755$$

4) คำนวณค่า IG ของแอตทริบิวต์ school_province “inCM” ต่อ GAT ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5197 \times \log_2 (0.5197) + 0.4803 \times \log_2 (0.4803)]$$

$$= 0.99888$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.25 \times \log_2 (0.25) + 0.75 \times \log_2 (0.75)]$$

$$= 0.81128$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6558 \times \log_2 (0.6558) + 0.3442 \times \log_2 (0.3442)]$$

$$= 0.92877$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4917 \times \log_2 (0.4917) + 0.5083 \times \log_2 (0.5083)]$$

$$= 0.9998$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.99888 - [0.002252 \times 0.81128 + 0.448198 \times 0.92877 + 0.54955 \times 0.9998]$$

$$= 0.99888 - [0.001827 \times 0.416272 \times 0.549441]$$

$$= 0.9946 - 0.9675$$

$$= 0.0313$$

5) คำนวณค่า IG ของแอตทริบิวต์ school_province “outCM” ต่อ school_size ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4576 \times \log_2 (0.4576) + 0.5424 \times \log_2 (0.5424)]$$

$$= 0.99481$$

$$\begin{aligned}
 \text{entropy (ผล=L)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.3269 \times \log_2 (0.3269) + 0.6731 \times \log_2 (0.6731)] \\
 &= 0.91175
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=M)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4885 \times \log_2 (0.4885) + 0.5115 \times \log_2 (0.5115)] \\
 &= 0.99962
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=S)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5672 \times \log_2 (0.5672) + 0.4328 \times \log_2 (0.4328)] \\
 &= 0.98694
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=XL)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4597 \times \log_2 (0.4597) + 0.5403 \times \log_2 (0.5403)] \\
 &= 0.99531
 \end{aligned}$$

$$\begin{aligned}
 \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล=L}) \times (\text{ผล=L}) + p(\text{ผล=M}) \times (\text{ผล=} \\
 &\text{M}) + p(\text{ผล=S}) \times (\text{ผล=S}) + p(\text{ผล=XL}) \times (\text{ผล=XL})] \\
 &= 0.99481 - [0.135593 \times 0.91175 + 0.226858 \times 0.99962 \\
 &+ 0.087353 \times 0.98694 + 0.550196 \times 0.99531] \\
 &= 0.99481 - [0.123627 + 0.226771 + 0.086213 + \\
 &0.547616] \\
 &= 0.99481 - 0.9842
 \end{aligned}$$

$$= 0.0106$$

6) คำนวณค่า IG ของแอตทริบิวต์ school_province “outCM” ต่อ gpax6_score ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4576 \times \log_2 (0.4576) + 0.5424 \times \log_2 (0.5424)]$$

$$= 0.99481$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4254 \times \log_2 (0.4254) + 0.5746 \times \log_2 (0.5746)]$$

$$= 0.98389$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.3725 \times \log_2 (0.3725) + 0.6275 \times \log_2 (0.6275)]$$

$$= 0.95261$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5028 \times \log_2 (0.5028) + 0.4972 \times \log_2 (0.4972)]$$

$$= 0.99998$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.99481 - [0.471969 \times 0.98389 + 0.066493 \times 0.066493 + 0.461538 \times 0.99998]$$

$$= 0.99481 - [0.288041 \times 0.120662 \times 0.564693]$$

$$= 0.99481 - 0.989234$$

$$= 0.005579$$

7) คำนวณค่า IG ของแอตทริบิวต์ school_province “outCM” ต่อ ONET ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4576 \times \log_2 (0.4576) + 0.5424 \times \log_2 (0.5424)]$$

$$= 0.99481$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4729 \times \log_2 (0.4729) + 0.5271 \times \log_2 (0.5271)]$$

$$= 0.99787$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4535 \times \log_2 (0.4535) + 0.5465 \times \log_2 (0.5465)]$$

$$= 0.99374$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล= Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})]$$

$$= 0.99481 - [0.336375 \times 0.99787 + 0.658409 \times 0.99374]$$

$$= 0.99481 - [0.335661 \times 0.65429]$$

$$= 0.99481 - 0.98995$$

$$= 0.004863$$

8) คำนวณค่า IG ของแอตทริบิวต์ school_province “outCM” ต่อ GAT ได้ดังนี้

$$\begin{aligned} \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4576 \times \log_2 (0.4576) + 0.5424 \times \log_2 (0.5424)] \\ &= 0.99481 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=High)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.1667 \times \log_2 (0.1667) + 0.8333 \times \log_2 (0.8333)] \\ &= 0.65002 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4699 \times \log_2 (0.4699) + 0.5301 \times \log_2 (0.5301)] \\ &= 0.99739 \end{aligned}$$

$$\begin{aligned} \text{entropy (ผล=Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\ &\log_2 p(\text{NotConfirm}) \\ &= -[0.4617 \times \log_2 (0.4617) + 0.5383 \times \log_2 (0.5383)] \\ &= 0.99576 \end{aligned}$$

$$\begin{aligned} \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=} \\ &\text{Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})] \\ &= 0.99481 - [0.023468 \times 0.65002 + 0.346806 \times \\ &0.99739 + 0.629726 \times 0.99576] \\ &= 0.99481 - [0.015255 \times 0.3459 \times 0.627058] \\ &= 0.99481 - 0.988213 \end{aligned}$$

$$= 0.0066$$

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ GAT มีค่ามากที่สุด (0.0313) ของโหนด school_province “inCM” และแอตทริบิวต์ GAT มีค่ามากที่สุด (0.0106) ของโหนด school_province “outCM” ดังนั้นจึงเลือกแอตทริบิวต์ school_size ขึ้นมาเป็นโหนดระดับ 3 และจะต้องทำการแตกกิ่งจากโหนดระดับ 4 ออกไปจนข้อมูลในแต่ละโหนดมีคลาสค่าตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 4 ต่อจากระดับที่ 3 เพื่อหาค่า IG ที่มากที่สุดของแอตทริบิวต์ GAT ต่อ school_size gpax6_score ONET และ GAT กับจำนวนข้อมูลที่ทั้งหมดโดยใช้ผลลัพธ์เป็น Confirm และ NotConfirm ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ school_size “M” ต่อ gpax6_score ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5016 \times \log_2 (0.5016) + 0.4984 \times \log_2 (0.4984)]$$

$$= 0.999992911$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5058 \times \log_2 (0.5058) + 0.4942 \times \log_2 (0.4942)]$$

$$= 0.999902466$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6296 \times \log_2 (0.6296) + 0.3704 \times \log_2 (0.3704)]$$

$$= 0.950956048$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4667 \times \log_2 (0.4667) + 0.5333 \times \log_2 (0.5333)]$$

$$= 0.996791632$$

$$IG (\text{parent, child}) = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.999992911 - [0.429824561 \times 0.999902466 + 0.087719298 \times 0.950956048 + 0.48245614 \times 0.996791632]$$

$$= 0.999992911 - [0.429783 \times 0.083417 \times 0.480908]$$

$$= 0.999992911 - 0.9941$$

$$= 0.0059$$

2) คำนวณค่า IG ของแอตทริบิวต์ school_size “M” ต่อ ONET ได้ดังนี้

$$\text{entropy} (\text{parent}) = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5016 \times \log_2 (0.5016) + 0.4984 \times \log_2 (0.4984)]$$

$$= 0.999992911$$

$$\text{entropy} (\text{ผล=High}) = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.1429 \times \log_2 (0.1429) + 0.8571 \times \log_2 (0.8571)]$$

$$= 0.591672779$$

$$\text{entropy} (\text{ผล=Low}) = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4954 \times \log_2 (0.4954) + 0.5046 \times \log_2 (0.5046)]$$

$$= 0.999939285$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5172 \times \log_2 (0.5172) + 0.4828 \times \log_2 (0.4828)]$$

$$= 0.999142104$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.999992911 - [0 \times 0.591672779 + 0.456140351 \times 0.999939285 + 0.543859649 \times 0.999142104]$$

$$= 0.999992911 - [0 \times 0.456113 \times 0.543393]$$

$$= 0.999992911 - 0.999506$$

$$= 0.000487$$

3) คำนวณค่า IG ของแอตทริบิวต์ school_size “M” ต่อ GAT ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5016 \times \log_2 (0.5016) + 0.4984 \times \log_2 (0.4984)]$$

$$= 0.999992911$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.3333 \times \log_2 (0.3333) + 0.6667 \times \log_2 (0.6667)]$$

$$= 0.918295834$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5333 \times \log_2 (0.5333) + 0.4667 \times \log_2 (0.4667)]$$

$$= 0.996791632$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4895 \times \log_2 (0.4895) + 0.5105 \times \log_2 (0.5105)]$$

$$= 0.999680266$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.999992911 - [0 \times 0.918295834 + 0.50877193 \times 0.996791632 + 0.49122807 \times 0.9996800266]$$

$$= 0.999992911 - [0 \times 0.50714 \times 0.491071]$$

$$= 0.999992911 - 0.998211$$

$$= 0.001782$$

4) คำนวณค่า IG ของแอตทริบิวต์ school_size “XL” ต่อ gpa6_score ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.50 \times \log_2 (0.50) + 0.50 \times \log_2 (0.50)]$$

$$= 1$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4625 \times \log_2 (0.4625) + 0.5375 \times \log_2 (0.5375)]$$

$$= 0.995938608$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.2759 \times \log_2(0.2759) + 0.7241 \times \log_2(0.7241)]$$

$$= 0.849751137$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.5629 \times \log_2(0.5629) + 0.4371 \times \log_2(0.4371)]$$

$$= 0.988548836$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 1 - [0.3077 \times 0.995938608 + 0.1115 \times 0.849751137 + 0.5808 \times 0.988548836]$$

$$= 1 - [0.306443 \times 0.09478 \times 0.574119]$$

$$= 1 - 0.9753$$

$$= 0.0247$$

5) คำนวณค่า IG ของแอดทริบิวต์ school_size “XL” ต่อ ONET ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.50 \times \log_2(0.50) + 0.50 \times \log_2(0.50)]$$

$$= 1$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4758 \times \log_2(0.4758) + 0.5242 \times \log_2(0.5242)]$$

$$= 0.998310441$$

$$\begin{aligned}
 \text{entropy (ผล=Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5221 \times \log_2 (0.5221) + 0.4779 \times \log_2 (0.4779)] \\
 &= 0.998595537
 \end{aligned}$$

$$\begin{aligned}
 \text{IG (parent, child)} &= \text{entropy}(\text{parent}) - [p(\text{ผล= Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})] \\
 &= 1 - [0.4769 \times 0.998310441 + 0.5231 \times 0.998595537] \\
 &= 1 - [0.476117 \times 0.522342] \\
 &= 1 - 0.99846 \\
 &= 0.00154
 \end{aligned}$$

6) คำนวณค่า IG ของแอตทริบิวต์ school_size “XL” ต่อ GAT ได้ดังนี้

$$\begin{aligned}
 \text{entropy (parent)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.50 \times \log_2 (0.50) + 0.50 \times \log_2 (0.50)] \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=Low)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.4661 \times \log_2 (0.4661) + 0.5339 \times \log_2 (0.5339)] \\
 &= 0.996681868
 \end{aligned}$$

$$\begin{aligned}
 \text{entropy (ผล=Moderate)} &= -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \\
 &\log_2 p(\text{NotConfirm}) \\
 &= -[0.5282 \times \log_2 (0.5282) + 0.4718 \times \log_2 (0.4718)]
 \end{aligned}$$

$$= 0.997709249$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล= Low}) \times (\text{ผล= Low}) + p(\text{ผล= Moderate}) \times (\text{ผล= Moderate})]$$

$$= 1 - [0.4538 \times 0.996681868 + 0.5462 \times 0.997709249]$$

$$= 1 - [0.45234 \times 0.544903]$$

$$= 1 - 0.997243$$

$$= 0.002757$$

7) คำนวณค่า IG ของแอตทริบิวต์ GAT “Moderate” ต่อ gpax6_score ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4216 \times \log_2 (0.4216) + 0.5784 \times \log_2 (0.5784)]$$

$$= 0.98221$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.3934 \times \log_2 (0.3934) + 0.6066 \times \log_2 (0.6066)]$$

$$= 0.966985$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6875 \times \log_2 (0.6875) + 0.3125 \times \log_2 (0.3125)]$$

$$= 0.896038$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4419 \times \log_2 (0.4419) + 0.5581 \times \log_2 (0.5581)]$$

$$= 0.990225$$

$$IG (\text{parent, child}) = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.98221 - [56903 \times 0.966985 + 0.029851 \times 0.896038 + 0.401119 \times 0.990225]$$

$$= 0.98221 - [550243 \times 0.026747 \times 0.397198]$$

$$= 0.98221 - 0.9742$$

$$= 0.00802$$

8) คำนวณค่า IG ของแอตทริบิวต์ GAT “Moderate” ต่อ gpax6_score ได้ดังนี้

$$\text{entropy} (\text{parent}) = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4216 \times \log_2 (0.4216) + 0.5784 \times \log_2 (0.5784)]$$

$$= 0.98221$$

$$\text{entropy} (\text{ผล=High}) = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.3934 \times \log_2 (0.3934) + 0.6066 \times \log_2 (0.6066)]$$

$$= 0.966985$$

$$\text{entropy} (\text{ผล=Low}) = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.6875 \times \log_2 (0.6875) + 0.3125 \times \log_2 (0.3125)]$$

$$= 0.896038$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4419 \times \log_2 (0.4419) + 0.5581 \times \log_2 (0.5581)]$$

$$= 0.990225$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.98221 - [56903 \times 0.966985 + 0.029851 \times 0.896038 + 0.401119 \times 0.990225]$$

$$= 0.98221 - [0.550243 \times 0.026747 \times 0.397198]$$

$$= 0.98221 - 0.9742$$

$$= 0.00802$$

9) คำนวณค่า IG ของแอดทริบิวต์ GAT “Moderate” ต่อ ONET ได้ดังนี้

$$\text{entropy (parent)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4216 \times \log_2 (0.4216) + 0.5784 \times \log_2 (0.5784)]$$

$$= 0.98221$$

$$\text{entropy (ผล=High)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.1429 \times \log_2 (0.1429) + 0.8571 \times \log_2 (0.8571)]$$

$$= 0.591673$$

$$\text{entropy (ผล=Low)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.3548 \times \log_2 (0.3548) + 0.6452 \times \log_2 (0.6452)]$$

$$= 0.938315$$

$$\text{entropy (ผล=Moderate)} = -p(\text{Confirm}) \times \log_2 p(\text{Confirm}) + p(\text{NotConfirm}) \times \log_2 p(\text{NotConfirm})$$

$$= -[0.4297 \times \log_2 (0.4297) + 0.5703 \times \log_2 (0.5703)]$$

$$= 0.9857$$

$$\text{IG (parent, child)} = \text{entropy}(\text{parent}) - [p(\text{ผล=High}) \times (\text{ผล=High}) + p(\text{ผล=Low}) \times (\text{ผล=Low}) + p(\text{ผล=Moderate}) \times (\text{ผล=Moderate})]$$

$$= 0.98221 - [0.1306 \times 0.591673 + 0.057836 \times 0.938315 + 0.929104 \times 0.9857]$$

$$= 0.98221 - [0.07727 \times 0.054268 \times 0.915819]$$

$$= 0.98221 - 0.977814$$

$$= 0.00439639$$

3.1.6 การสร้างโมเดล Decision Tree ด้วยโปรแกรม Weka

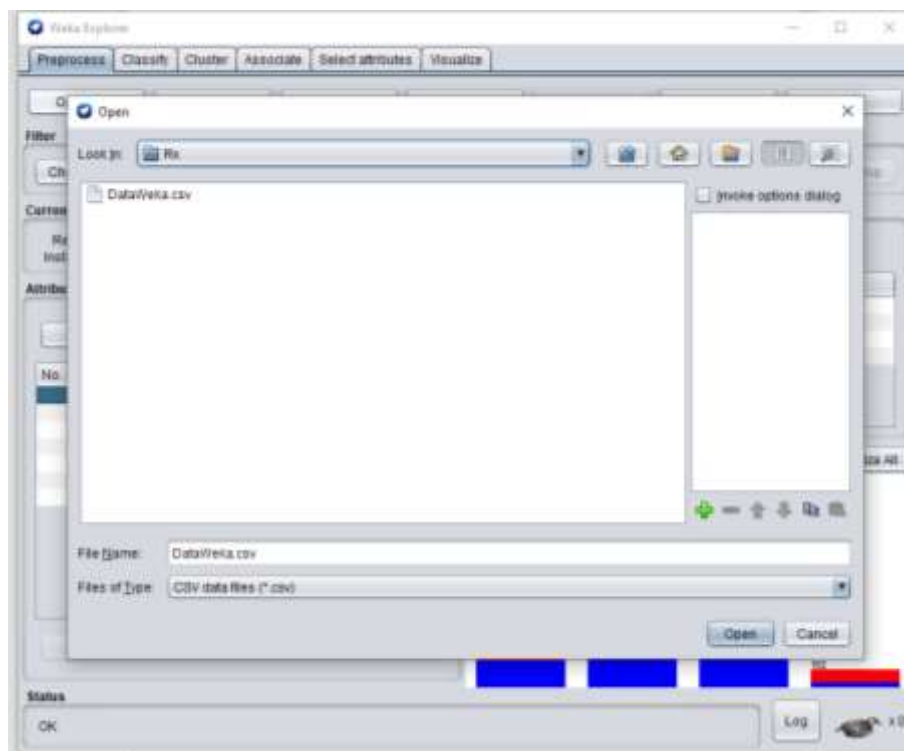
คณะผู้จัดทำได้ทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self-Consistency Test เหมาะสำหรับการทดสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือก มาทดสอบกับข้อมูลที่ผ่านมาการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าวมาทดสอบกับโปรแกรม Weka และโปรแกรม Knime ซึ่งมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 คลิกเปิดโปรแกรม Weka

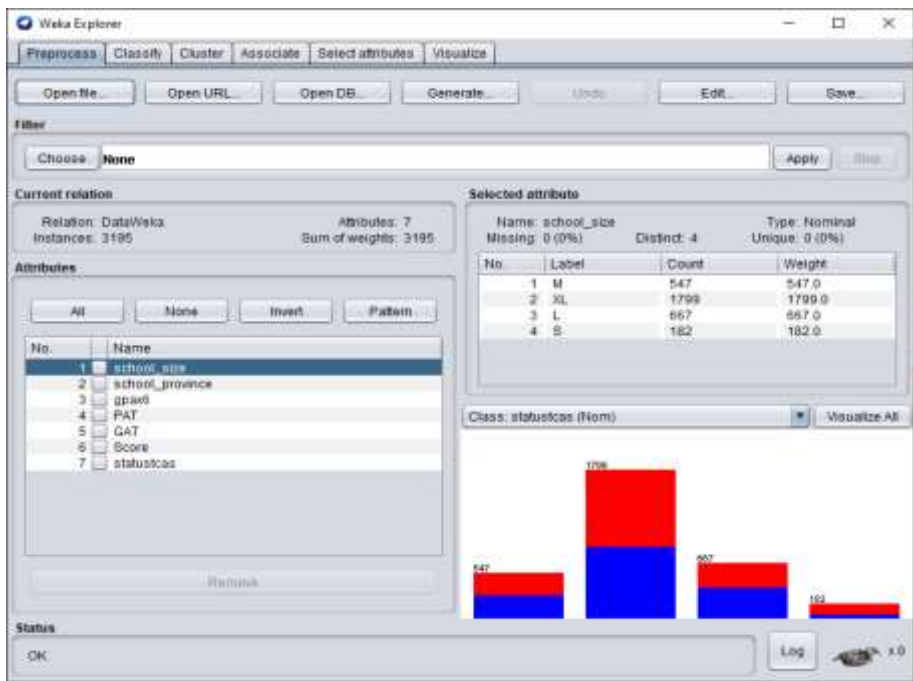


ภาพที่ 3.10 แสดงการเข้าหน้าจอโปรแกรม Weka

ขั้นตอนที่ 2 นำข้อมูลที่จัดเตรียมไว้ โดยเลือกที่ Application >> Explore
>>Openfile เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบ

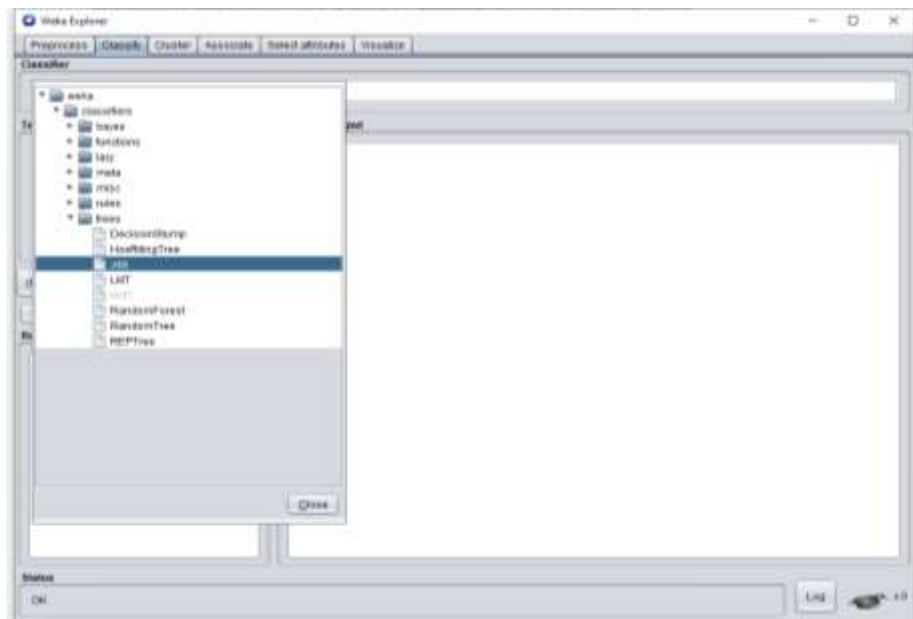


ภาพที่ 3.11 แสดงการนำข้อมูลเข้าในโปรแกรม Weka



ภาพที่ 3.12 หน้าจอแสดงข้อมูลที่นำเข้าไปรแกรม

ขั้นตอนที่ 3 คลิกที่แท็บ Classify หน้าตาของ workspace จะเปลี่ยนไปเป็นส่วน ของแท็บ Classify แล้วคลิกที่ปุ่ม Choose จะมีลิสต์ (list) แสดงเทคนิคต่างๆของการ Classify ให้คลิกเลือกที่เมนู Trees หลังจากนั้นให้เลือกที่เทคนิคการจำแนกข้อมูลด้วย J48 โดยคลิกที่เมนู J48 หลังจากนั้นคลิกที่ปุ่ม Start Weka จะทำงานและแสดงผลลัพธ์การทำงาน



ภาพที่ 3.13 การเลือกเทคนิคที่ใช้ในการจำแนกข้อมูลแบบ Decision Tree: J48

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2526      79.061 %
Incorrectly Classified Instances    669       20.939 %
Kappa statistic                     0.8008
Mean absolute error                 0.3218
Root mean squared error             0.4025
Relative absolute error             64.3895 %
Root relative squared error        80.5152 %
Total Number of Instances          3195

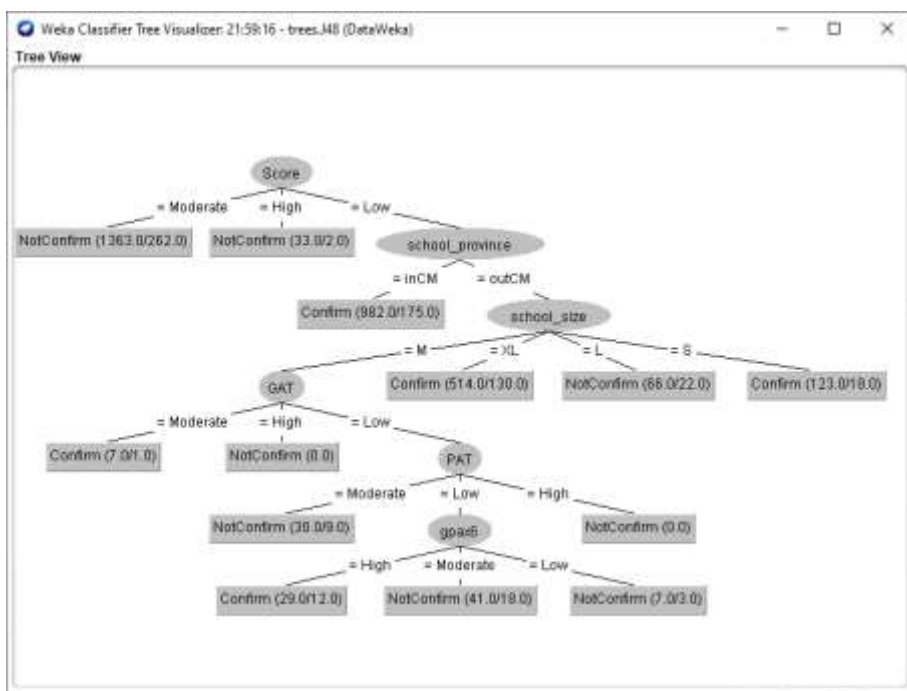
=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.778   0.197   0.790   0.778   0.784   0.881  0.802   0.755   NotConfirm
0.803   0.222   0.791   0.803   0.797   0.881  0.802   0.770   Confirm
Weighted Avg.  0.791   0.210   0.791   0.791   0.791   0.581  0.802   0.763

=== Confusion Matrix ===

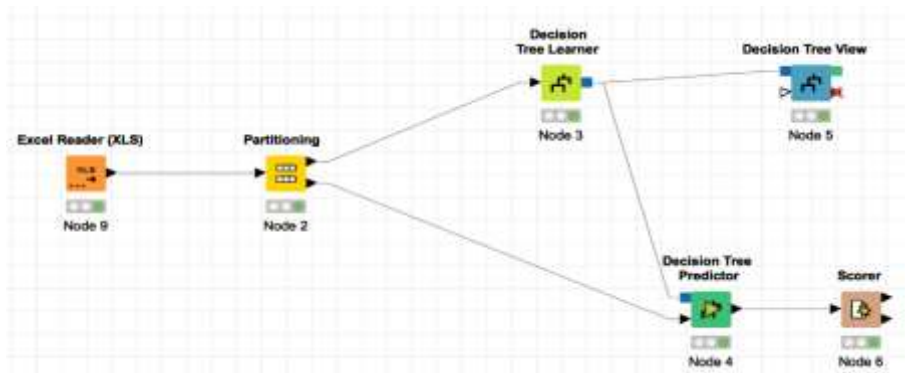
  a  b  <-- classified as
1213 347 |  a = NotConfirm
 322 1313 |  b = Confirm
    
```

ภาพที่ 3.14 ผลการจำแนกกลุ่มแบบ Decision Tree: J48 ในโปรแกรม Weka



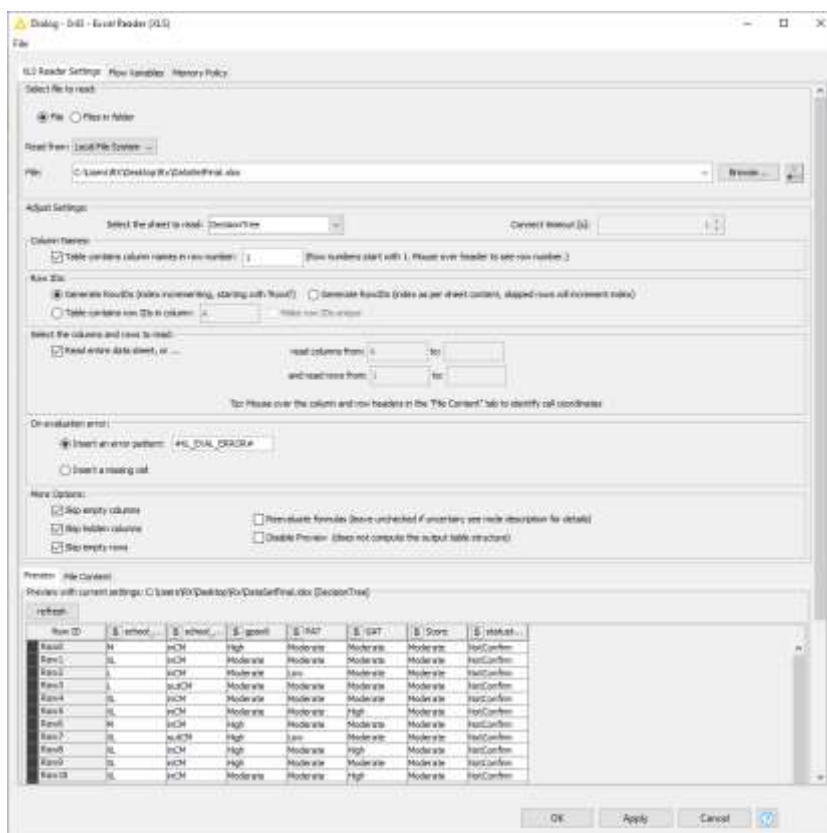
ภาพที่ 3.15 ผล Visualization Tree จากการวิเคราะห์ข้อมูลในโปรแกรม Weka

3.1.7 การสร้างโมเดล Decision Tree ด้วยโปรแกรม Knime



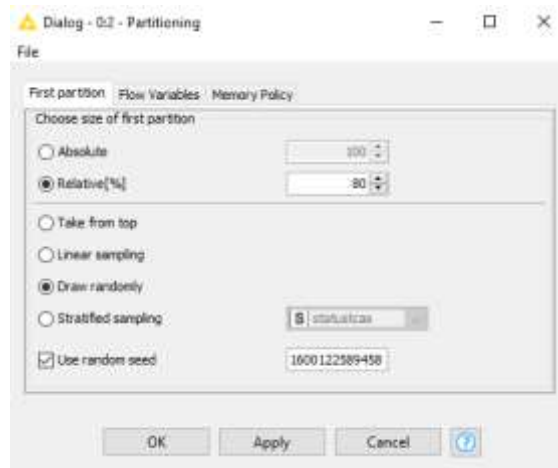
ภาพที่ 3.16 กระบวนการวิเคราะห์ข้อมูลโปรแกรม Knime

Read an XLS file โหนดนี้หลักการทำงานอ่านสเปรดชีตและจัดเตรียมไว้ที่พอร์ตเอาต์พุตอ่านเฉพาะข้อมูล แต่สามารถอ่านไฟล์หลายไฟล์ที่มีชื่อแผ่นงานเดียวกันพร้อมกันได้สามารถอ่านได้เฉพาะข้อมูลได้แก่ String, Time AndDate, Double, Boolean และ Int



ภาพที่ 3.17 แสดงการนำข้อมูลเข้าโปรแกรม Knime

Partitioning ตารางอินพุตแบ่งออกเป็นสองพาร์ติชัน (เช่น row-wise) เช่นฝึกอบรมหรือเรียนรู้(Training) และทดสอบข้อมูล(Testing) สองพาร์ติชันมีอยู่ที่พอร์ตเอาต์พุตสองพอร์ต



ภาพที่ 3.18 แสดงหน้าอินพุตแบ่งออกเป็นสองพาร์ติชัน

Decision Tree Learner โหนดนี้ทำให้เกิดแผนภูมิการตัดสินใจการจำแนกประเภทในหน่วยความจำหลัก เพื่อส่งแอตทริบิวต์ให้เชื่อมต่อกับพอร์ต Model out กับโหนด "Decision Tree Predictor"

Decision Tree Predictor โหนดนี้ใช้แผนผังการตัดสินใจที่มีอยู่ (ส่งผ่านพอร์ตโมเดล) เพื่อทำนายค่าคลาสสำหรับรูปแบบใหม่

Classified Data - 04 - Decision Tree Predictor

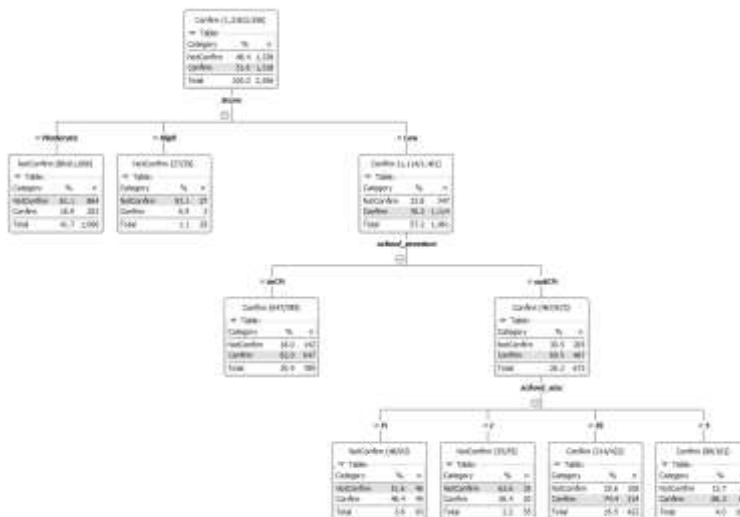
File Edit Help Navigation View

Table 'C:\Users\FX\Downloads\DT-dataset.xlsx [DTInme]' - Rows: 335 Spec - Columns: 8 Properties Flow Variables

Row ID	S school...	S school...	S spend...	S PAT	S GAT	S Score	S statat...	S Predict...
Row8	XL	inCM	Moderate	Moderate	Moderate	Moderate	NetConfirm	NetConfirm
Row23	XL	outCM	Low	Low	Low	Low	NetConfirm	NetConfirm
Row26	XL	inCM	High	Low	Moderate	Low	Confirm	Confirm
Row27	XL	inCM	Moderate	Low	Low	Low	NetConfirm	Confirm
Row28	XL	inCM	High	Low	Low	Low	Confirm	Confirm
Row33	XL	inCM	High	Moderate	High	Moderate	NetConfirm	NotConfirm
Row37	XL	outCM	Moderate	Low	Low	Low	NetConfirm	Confirm
Row46	M	inCM	High	Low	Low	Low	NetConfirm	Confirm
Row50	XL	outCM	Moderate	Moderate	Moderate	Moderate	NetConfirm	NotConfirm
Row51	M	inCM	Moderate	Low	Low	Low	NetConfirm	Confirm
Row54	XL	outCM	Moderate	Low	Low	Moderate	Confirm	NotConfirm
Row58	L	outCM	Moderate	Moderate	Low	Low	Confirm	NotConfirm
Row61	XL	outCM	High	Moderate	Low	Moderate	NetConfirm	NotConfirm
Row77	XL	outCM	High	Moderate	Low	Moderate	NetConfirm	NotConfirm
Row79	XL	outCM	High	Low	Moderate	Low	NetConfirm	NotConfirm
Row88	L	inCM	High	Moderate	Low	Moderate	Confirm	NotConfirm
Row99	L	inCM	Moderate	Moderate	Moderate	Moderate	NetConfirm	NotConfirm
Row102	XL	outCM	Moderate	Low	Low	Low	NetConfirm	Confirm
Row109	L	inCM	High	Moderate	Low	Moderate	NetConfirm	NotConfirm
Row114	L	outCM	High	Moderate	Low	Moderate	NetConfirm	NotConfirm
Row126	M	outCM	Moderate	Low	Low	Low	NetConfirm	NotConfirm
Row129	M	outCM	High	Low	Low	Low	Confirm	NotConfirm
Row128	L	outCM	High	Moderate	Low	Low	NetConfirm	NotConfirm
Row135	XL	inCM	High	Low	Low	Low	NetConfirm	Confirm
Row139	M	outCM	Moderate	Low	Low	Low	NetConfirm	NotConfirm

ภาพที่ 3.19 Decision Tree Predictor เพื่อทำนายค่าคลาสสำหรับรูปแบบใหม่

Decision Tree View มุมมองต้นไม้ตัดสินใจแผนผังการตัดสินใจตามทีละบัพในพอร์ตโมเดลพร้อมกับข้อมูลที่จัดประเภทต้นไม้สามารถขยายและยุบได้โดยใช้เครื่องหมายบวก / ลบ



ภาพที่ 3.20 Decision Tree View แผนผังการตัดสินใจ

Scorer เปรียบเทียบสองคอลัมน์ตามคู่ค่าแอตทริบิวต์และแสดง Confusion Matrixคือจำนวนแถวที่แอตทริบิวต์และการจัดหมวดหมู่ตรงกัน

Row ID	TruePo...	FalsePo...	TrueNe...	FalseNe...	D Recall	D Precision	D Sensitivity	D Specificity	D F-meas...	D Accuracy	D Coher...
NotConfirm	263	70	247	59	0.817	0.79	0.817	0.779	0.803		
Confirm	247	59	263	70	0.779	0.807	0.779	0.817	0.793		
Overall										0.798	0.596

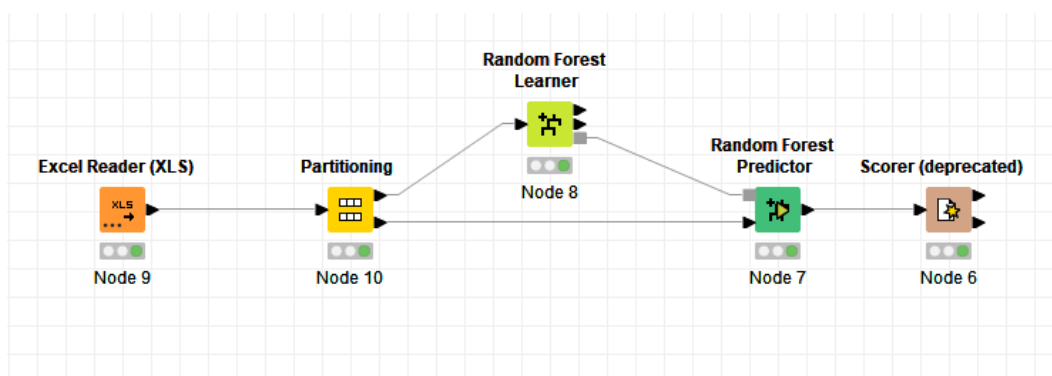
ภาพที่ 3.21 ผลการคำนวณความแม่นยำของข้อมูล (Accuracy Statistics)

statustcas ...	NotConfirm	Confirm
NotConfirm	263	59
Confirm	70	247

Correct classified: 510 Wrong classified: 129
 Accuracy: 79.812 % Error: 20.188 %
 Cohen's kappa (κ) 0.596

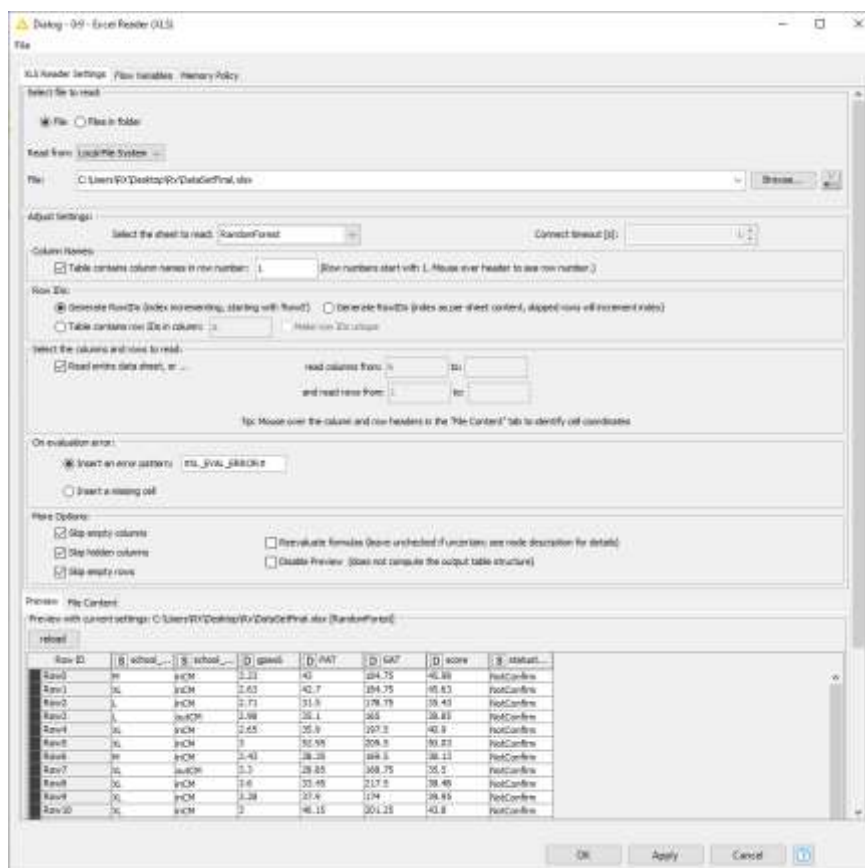
ภาพที่ 3.22 ผลการคำนวณคอนฟิวชั่น เมทริกซ์ (Confusion Matrix)

3.1.7 การสร้างโมเดล Random Forest ด้วยโปรแกรม Knime



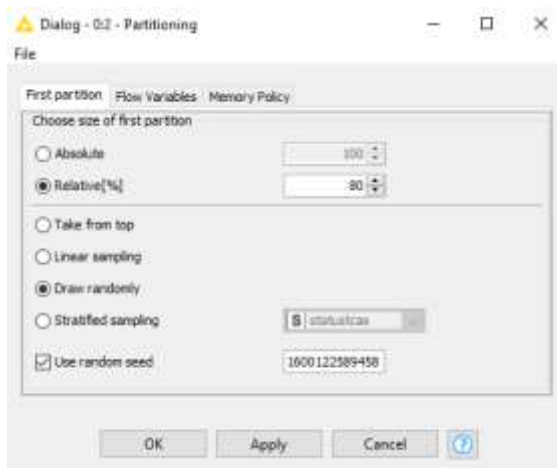
ภาพที่ 3.23 กระบวนการวิเคราะห์การจำแนกข้อมูลแบบ Random Forest โปรแกรม Knime

Read an XLS file โหนดนี้หลักการทำงานอ่านสเปรดชีทและจัดเตรียมไว้ที่พอร์ตเอาต์พุตอ่านเฉพาะข้อมูล แต่สามารถอ่านไฟล์หลายไฟล์ที่มีชื่อแผ่นงานเดียวกันพร้อมกันได้สามารถอ่านได้เฉพาะข้อมูลได้แก่ String, Time AndDate, Double, Boolean และ Int



ภาพที่ 3.24 แสดงการนำข้อมูลเข้าในโปรแกรม Knime

Partitioning ตารางอินพุตแบ่งออกเป็นสองพาร์ติชัน (เช่น row-wise) เช่นฝึกอบรมหรือเรียนรู้(Training) และทดสอบข้อมูล(Testing) สองพาร์ติชันมีอยู่ที่พอร์ตเอาต์พุตสองพอร์ต



ภาพที่ 3.25 แสดงหน้าอินพุตแบ่งออกเป็นสองพาร์ติชัน

Random Forest Learner โหนดนี้จัดเตรียมชุดย่อยของฟังก์ชันการทำงานของ Tree Ensemble Learner ที่สอดคล้องกับฟอเรสต์แบบสุ่ม

Random Forest Predictor โหนดนี้ใช้ทำนายรูปแบบตามการรวมการคาดการณ์ของต้นไม้แต่ละต้นในแบบจำลองฟอเรสต์แบบสุ่ม

Row ID	school	gender	race	math	reading	science	status	predicted	predicted error
Row02	M	MCH	L	43	34.72	45.98	Not Confirmed	Not Confirmed	0.93
Row03	L	notCM	L	75.1	189	98.69	Not Confirmed	Not Confirmed	0.92
Row07	M	notCM	L	29.89	198.79	58.9	Not Confirmed	Not Confirmed	1
Row08	M	MCH	L	45.25	25.23	42.8	Not Confirmed	Not Confirmed	0.81
Row11	M	notCM	L	37.27	185	38.94	Not Confirmed	Not Confirmed	0.88
Row18	M	notCM	L	45.2	133	38.7	Not Confirmed	Not Confirmed	1
Row17	L	MCH	L	83.99	187.5	97.84	Not Confirmed	Not Confirmed	0.94
Row19	M	MCH	L	88.8	188	91.38	Not Confirmed	Not Confirmed	1
Row20	M	notCM	L	46	128.35	38.67	Not Confirmed	Not Confirmed	1
Row23	M	notCM	L	25.1	81.28	38.47	Not Confirmed	Not Confirmed	0.97
Row24	M	MCH	L	88.9	188	91.38	Not Confirmed	Not Confirmed	1
Row32	L	MCH	L	25.34	0	18.44	Not Confirmed	Not Confirmed	0.98
Row35	M	notCM	L	26.9	265.78	42.68	Not Confirmed	Not Confirmed	0.89
Row38	M	MCH	L	25.35	157	38.28	Not Confirmed	Not Confirmed	0.87
Row43	M	MCH	L	26.7	0	28.12	Not Confirmed	Not Confirmed	0.78
Row42	M	MCH	L	30.9	138	142.88	Not Confirmed	Not Confirmed	0.98
Row44	M	MCH	L	31.16	178.5	152.82	Not Confirmed	Not Confirmed	0.82
Row45	M	MCH	L	45.72	181.25	42.38	Not Confirmed	Not Confirmed	0.89
Row48	M	MCH	L	26.2	200	41.9	Not Confirmed	Not Confirmed	0.82
Row49	M	notCM	L	26.09	177.9	37.38	Not Confirmed	Not Confirmed	0.86
Row49	M	notCM	L	26.29	187	42.38	Not Confirmed	Not Confirmed	0.76
Row50	L	MCH	L	17	157.5	38.45	Not Confirmed	Not Confirmed	0.82
Row54	M	notCM	L	40.8	181.25	42.63	Not Confirmed	Not Confirmed	0.83
Row56	M	MCH	L	29.09	235	42.03	Not Confirmed	Not Confirmed	0.94
Row60	L	MCH	L	38.3	168.28	38.38	Not Confirmed	Not Confirmed	0.83
Row60	M	notCM	L	24.25	120	48.48	Not Confirmed	Not Confirmed	0.88
Row62	L	MCH	L	38.5	205.72	42.63	Not Confirmed	Not Confirmed	0.99
Row65	M	MCH	L	45.8	177	42.5	Not Confirmed	Not Confirmed	0.88
Row104	M	MCH	L	48.9	229.8	96.7	Not Confirmed	Not Confirmed	1
Row108	M	notCM	L	34.8	178	68.1	Not Confirmed	Not Confirmed	0.83
Row114	M	MCH	L	41.45	157.5	38.68	Not Confirmed	Not Confirmed	0.83
Row117	L	notCM	L	38.8	170.25	38.55	Not Confirmed	Not Confirmed	1
Row122	L	MCH	L	37.30	140	38.18	Not Confirmed	Not Confirmed	0.97
Row128	L	MCH	L	38.48	187.29	37.18	Not Confirmed	Not Confirmed	0.88
Row131	M	notCM	L	31.81	236.25	48.78	Not Confirmed	Not Confirmed	0.82
Row137	M	notCM	L	25.3	152.3	48.88	Not Confirmed	Not Confirmed	0.82
Row140	L	MCH	L	37	182.25	37.28	Not Confirmed	Not Confirmed	0.88
Row144	M	notCM	L	48	189.78	38.12	Not Confirmed	Not Confirmed	1

ภาพที่ 3.26 Decision Tree Predictor เพื่อทำนายค่าคลาสสำหรับรูปแบบใหม่

Scorer เปรียบเทียบสองคอลัมน์ตามค่าแอตทริบิวต์และแสดง Confusion Matrix คือจำนวนแถวที่แอตทริบิวต์และการจัดหมวดหมู่ตรงกัน

Row ID	TruePos...	FalsePo...	TrueFte...	FalseF...	D Recall	D Precision	D Sensitivity	D Specificity	D F-meas...	D Accuracy	D Cohen...
NotConfirm	266	58	256	59	0.818	0.821	0.818	0.815	0.82	?	?
Confirm	256	59	266	58	0.815	0.813	0.815	0.818	0.814	?	?
Overall	?	?	?	?	?	?	?	?	?	0.817	0.634

ภาพที่ 3.27 ผลการคำนวณความแม่นยำของข้อมูล (Accuracy Statistics)

statustcas ...	NotConfirm	Confirm
NotConfirm	266	59
Confirm	58	256

Correct classified: 522 Wrong classified: 117
 Accuracy: 81.69 % Error: 18.31 %
 Cohen's kappa (κ) 0.634

ภาพที่ 3.28 ผลการคำนวณคอนฟิวชั่น เมทริกซ์ (Confusion Matrix)

3.1.8 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งานด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรกว่า มีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

ภาพที่ 3.29 คอนฟิวชั่น เมทริกซ์ (Confusion Matrix)

ที่มา : medium.com (2562)

True Positive (TP)= สิ่งที่ทำนาย ตรงกับสิ่งที่เกิดขึ้นจริง ในกรณี ทำนายว่าจริง และสิ่งที่เกิดขึ้นก็คือ จริง

กรณีนี้คือ ทำนายว่านักเรียนยืนยันสิทธิ์และนักเรียนยืนยันสิทธิ์จริง

True Negative (TN)= สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น ในกรณีทำนายว่า ไม่จริง และสิ่งที่เกิดขึ้นก็คือ ไม่จริง

กรณีนี้คือ ทำนายว่านักเรียนไม่ยืนยันสิทธิ์และนักเรียนไม่ยืนยันสิทธิ์จริง

False Positive (FP)= สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น คือทำนายว่า จริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง

กรณีนี้คือ ทำนายว่านักเรียนยืนยันสิทธิ์และนักเรียนไม่ยืนยันสิทธิ์

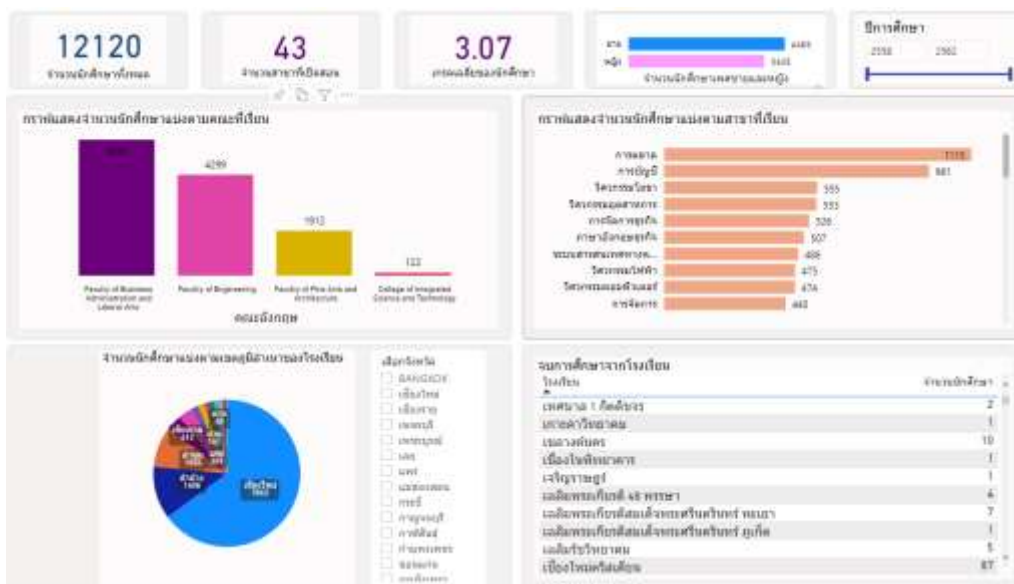
False Negative (FN)= สิ่งที่ทำนายไม่ตรงกับที่ที่เกิดขึ้นจริง คือทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้น คือ จริง

กรณีนี้คือ ทำนายว่านักเรียนไม่ยืนยันสิทธิ์และนักเรียนยืนยันสิทธิ์

โดย TP,TN,FP,FN ในตารางจะแทนด้วยค่าความถี่

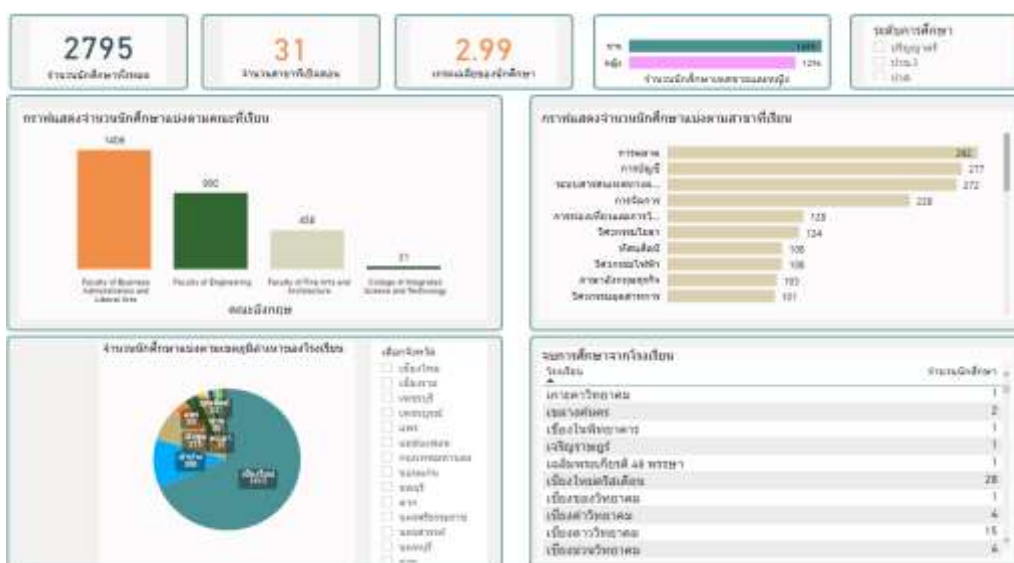
3.1.9 เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อม นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจ คณะผู้จัดทำนำผลข้อมูลที่ทำกรวิเคราะห์แล้ว แสดงผลข้อมูลบน Web browser โดยใช้ชุดคำสั่ง HTML และ CSS3 ด้วยการนำเสนอข้อมูลแบบ Visualization เป็นการแสดงผล ข้อมูลในรูปแบบของภาพ โดยใช้โปรแกรม Power Bi ซึ่งทางคณะผู้จัดทำได้ยกตัวอย่างการ จัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) กับข้อมูลนักเรียนย้อนหลังปี 2558-2562 โดยแต่ละปีจะมีรูปแบบเดียวกัน ดังนี้

1) แดชบอร์ดของข้อมูลนักศึกษาที่เข้ามาศึกษาปี 2558-2562 แสดงผลโดยแยกเป็นเพศ เกรดเฉลี่ย คณะที่เรียน สาขาที่เรียน ภูมิภาคของโรงเรียน จำนวนของนักศึกษา ผู้ใช้สามารถกรองดูข้อมูลได้จากตัวกรอง



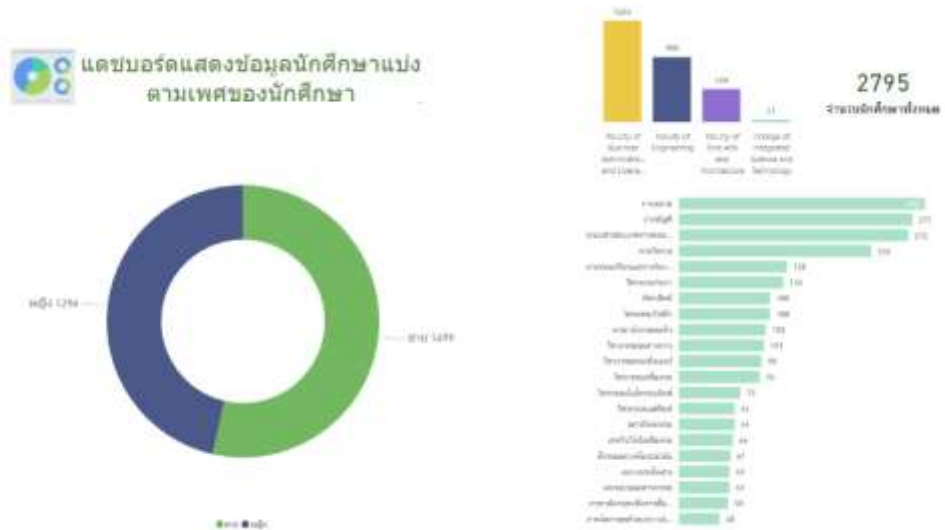
ภาพที่ 3.30 แดชบอร์ดนักศึกษาที่เข้ามาศึกษาปี 2558-2562

2) แดชบอร์ดของข้อมูลนักศึกษาที่เข้ามาศึกษาปีในแต่ละปีการศึกษา แสดงผลโดยรวมเป็นเพศ เกรดเฉลี่ย คณะที่เรียน สาขาที่เรียน ภูมิภาคของโรงเรียน จำนวนของนักศึกษาทั้งหมด



ภาพที่ 3.31 แดชบอร์ดข้อมูลนักศึกษาที่เข้ามาศึกษาในแต่ละปีการศึกษา

5) แดชบอร์ดแสดงข้อมูลจำนวนของนักศึกษาแบ่งตามเพศ แสดงผลโดยรวมเป็นกราฟ จำนวนนักศึกษาที่ได้เข้ามาศึกษาทั้งหมด จำนวนนักศึกษาตามคณะที่เรียน สาขาที่เรียน และตามเพศ



ภาพที่ 3.34 แดชบอร์ดข้อมูลนักศึกษาที่เข้ามาศึกษาในแต่ละปีการศึกษาแบ่งตามเพศ

6) แดชบอร์ดข้อมูลนักศึกษาที่เข้ามาศึกษาในแต่ละปีการศึกษาแบ่งคณะที่เรียน แสดงผลโดยรวมเป็นกราฟจำนวนนักศึกษาที่ได้เข้ามาศึกษาทั้งหมด เกรดเฉลี่ยของนักศึกษา จำนวนนักศึกษาแบ่งตามคณะที่เรียน และตารางจำนวนนักศึกษาแบ่งตามสาขาที่เรียน



ภาพที่ 3.35 แผนภาพวงกลมแสดงผลข้อมูลของนักศึกษาแบ่งตามสาขาและคณะที่เรียน

3.2 การออกแบบเว็บไซต์

3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

- 1) หน้าของเว็บไซต์ หน้าแสดงผลแดชบอร์ดสรุปข้อมูลนักศึกษาย้อนหลังปีการศึกษา 2558-2562



ภาพที่ 3.36 Wireframe หน้าแสดงผลแดชบอร์ดสรุปข้อมูลย้อนหลังปี 2558-2562

- 2) หน้าแสดงขั้นตอนการวิเคราะห์ข้อมูลที่จะมีการแสดงขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูลขั้นตอนที่ 1



ภาพที่ 3.37 Wireframe ขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูลขั้นตอนที่ 1

3) หน้าแสดงขั้นตอนการวิเคราะห์ข้อมูลที่จะมีการแสดงขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูลขั้นตอนที่ 2



ภาพที่ 3.38 Wireframe ขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูลขั้นตอนที่ 2

4) หน้าแสดงขั้นตอนการวิเคราะห์ข้อมูลที่จะมีการแสดงขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูลขั้นตอนที่ 3



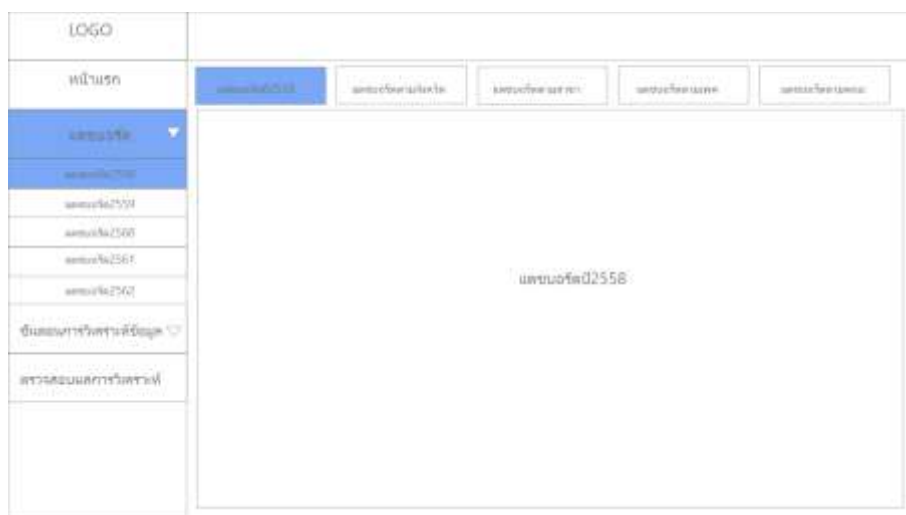
ภาพที่ 3.39 Wireframe ขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูล ขั้นตอนที่ 3

5) หน้าแสดงขั้นตอนการวิเคราะห์ข้อมูลที่จะมีการแสดงขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูลขั้นตอนที่ 4



ภาพที่ 3.40 Wireframe ขั้นตอนการสร้างโมเดลและวิเคราะห์ข้อมูล ขั้นตอนที่ 4

6) หน้าแสดงแดชบอร์ดสรุปข้อมูลนักศึกษาย้อนหลังแต่ปี



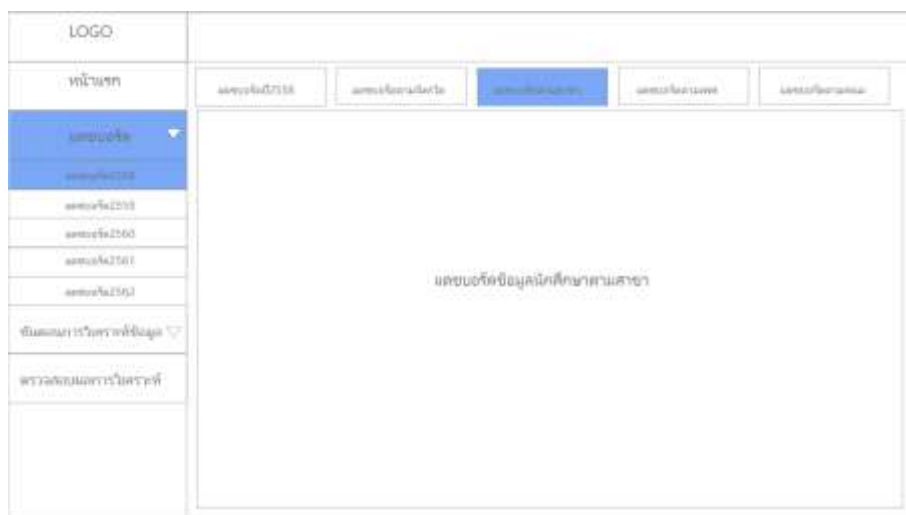
ภาพที่ 3.41 Wireframe หน้าแสดงแดชบอร์ดสรุปข้อมูลนักศึกษาย้อนหลังแต่ปี

7) หน้าแสดงแดชบอร์ดข้อมูลนักศึกษาแบ่งตามจังหวัดย้อนหลังแต่ละปี



ภาพที่ 3.42 Wireframe หน้าแสดงแดชบอร์ดข้อมูลนักศึกษาแบ่งตามจังหวัดย้อนหลังแต่ละปี

8) หน้าแสดงผลแดชบอร์ดข้อมูลนักศึกษาแบ่งตามสาขาย้อนหลังแต่ละปี



ภาพที่ 3.43 Wireframe หน้าแสดงแดชบอร์ดข้อมูลนักศึกษาแบ่งตามสาขาย้อนหลังแต่ละปี

9) หน้าแสดงผลแดชบอร์ดข้อมูลนักศึกษาแบ่งตามเพศย้อนหลังแต่ละปี

LOGO					
หน้าแรก	แดชบอร์ด2558	แดชบอร์ด2559	แดชบอร์ด2560	แดชบอร์ด2561	แดชบอร์ด2562
แดชบอร์ด	แดชบอร์ดข้อมูลนักศึกษาตามเพศ				
แดชบอร์ด2558					
แดชบอร์ด2559					
แดชบอร์ด2560					
แดชบอร์ด2561					
แดชบอร์ด2562					
แจ้งเตือนการชำระเงิน					
ตรวจสอบผลการเรียน					

ภาพที่ 3.44 Wireframe หน้าแสดงผลแดชบอร์ดข้อมูลนักศึกษาแบ่งตามเพศย้อนหลังแต่ละปี

10) หน้าแสดงผลแดชบอร์ดข้อมูลนักศึกษาแบ่งตามคณะย้อนหลังแต่ละปี

LOGO					
หน้าแรก	แดชบอร์ด2558	แดชบอร์ด2559	แดชบอร์ด2560	แดชบอร์ด2561	แดชบอร์ด2562
แดชบอร์ด	แดชบอร์ดข้อมูลนักศึกษาตามคณะ				
แดชบอร์ด2558					
แดชบอร์ด2559					
แดชบอร์ด2560					
แดชบอร์ด2561					
แดชบอร์ด2562					
แจ้งเตือนการชำระเงิน					
ตรวจสอบผลการเรียน					

ภาพที่ 3.45 Wireframe หน้าแสดงผลแดชบอร์ดข้อมูลนักศึกษาแบ่งตามคณะย้อนหลังแต่ละปี

11) หน้าการสร้าง Decition Tree บนเว็บไซต์โดยการนำเข้าไฟล์CSVและทำการประมวลผลจะได้รูปแบบ Decition Tree แบบเป็น Level



ภาพที่ 3.46 Wireframe หน้าการสร้าง Decition Tree บนเว็บไซต์

3.3 บทสรุป

วัตถุประสงค์เพื่อวิเคราะห์ข้อมูลผู้สมัครและข้อมูลนักศึกษาเพื่อสนับสนุนการตัดสินใจในการวางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS ด้วยการนำเทคนิคเหมืองข้อมูล (Data mining) การจำแนกข้อมูล (Data classification) และรูปแบบของแผนภูมิต้นไม้ (Decision tree) เปรียบเทียบกับวิธีการจำแนกข้อมูลแบบ Random Forest เพื่อวิเคราะห์หารูปแบบที่เกิดขึ้นจากข้อมูลการยื่นยันสิทธิ์เคลียร์ริงเฮาส์ (Clearing house) สามารถนำรูปแบบ (Model) ที่ได้มาวางแผนการคัดเลือกบุคคล เข้าศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา นำไปสู่การวางแผนประชาสัมพันธ์ต่อไปในอนาคต เป็นขั้นตอนแรกของกระบวนการนำข้อมูลรับสมัครนักศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ปีการศึกษา 2563 จำนวนรายการทั้งหมด 3,195 รายการ รวบรวมข้อมูลและพิจารณาตรวจสอบความถูกต้องและความครบถ้วนของข้อมูล พิจารณาข้อมูลทั้งหมดว่าส่วนใดของข้อมูลที่จำเป็นต่อการวิเคราะห์ข้อมูลให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้ นำข้อมูลที่ได้มาทำการคัดเลือกข้อมูล (Data Selection) การกลั่นกรองข้อมูล (Data Cleaning) การแปลงรูปแบบของข้อมูล (Data Transformation) ผู้จัดทำวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification ด้วยการสร้างโมเดล Decision Tree และโมเดล Random Forest ทำการทดสอบโมเดลเพื่อวัดประสิทธิภาพด้วยวิธี Self Consistency Test เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้นด้วยโปรแกรม knime ผลการวิเคราะห์สรุปได้ว่าโมเดล Random Forest เมื่อทำการทดสอบโมเดลเพื่อวัดประสิทธิภาพด้วยวิธี Self Consistency Test แล้วพบว่าค่าความแม่นยำ (Accuracy Statistics) มีความแม่นยำถึง 81.69% ซึ่งมากกว่า โมเดล Decision Tree ที่มีค่าความแม่นยำ 79.81% จัดทำเว็บไซต์เพื่อแสดงผลประกอบด้วย 4 ส่วนคือ

- 1) หน้าแรกของเว็บไซต์แสดงแดชบอร์ดข้อมูลนักศึกษาโดยรวมของทุกปี
- 2) แดชบอร์ดตามปีการศึกษาตั้งแต่ 2558-2562
 - หน้าแดชบอร์ดแสดงแดชบอร์ดข้อมูลนักศึกษาโดยรวมของปีทีเลือก
 - หน้าแดชบอร์ดแสดงข้อมูลนักศึกษาแบ่งตามภูมิภาค
 - หน้าแดชบอร์ดแสดงข้อมูลนักศึกษาแบ่งตามคณะที่เรียน

- หน้าแดชบอร์ดแสดงข้อมูลนักศึกษาแบ่งตามเพศ
- หน้าแดชบอร์ดแสดงข้อมูลนักศึกษาแบ่งตามระดับการศึกษา

3) ขั้นตอนการวิเคราะห์ข้อมูล

- หน้าแสดงข้อมูลการทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)
- หน้าแสดงข้อมูลเกี่ยวกับขั้นตอนการเตรียมข้อมูล (Data Preparation)
- หน้าแสดงขั้นตอนการตัดสินใจแบบโครงสร้างต้นไม้ (Decision Tree Weka)

4) ฟังก์ชันสร้างโมเดล Decision Tree