

DSI@TU

Data Science & Innovation
Thammasat University

รายงานฉบับสมบูรณ์

การวิเคราะห์ปัจจัยที่ทำให้เด็กไทยหลุดออกจากระบบการศึกษา และแนวทางการแก้ไข

โดย

- | | |
|-----------------------------|-----------------------|
| 1. นายดุสิตภาพ โตวิวัฒน์ | เลขทะเบียน 6309658018 |
| 2. นางสาวธนัชพร ใบสมุทร | เลขทะเบียน 6309658034 |
| 3. นางสาววรรณเวรา แยมบางยาง | เลขทะเบียน 6309658042 |
| 4. นางสาวโชติมา ธนอสนีกร | เลขทะเบียน 6309658059 |
| 5. นางสาวสุภาพสร สัมภ์แก้ว | เลขทะเบียน 6309658067 |

อาจารย์ที่ปรึกษาโครงการ: ผู้ช่วยศาสตราจารย์ ดร. รัชฎา คงคะจันทร์

รายงานนี้เป็นส่วนหนึ่งของรายวิชา วช314 โครงการด้านการวิเคราะห์ข้อมูลสำหรับธุรกิจ (DSI314
Business Analytics Capstone Project)

ภาคการศึกษา 1/2565

สารบัญ

สารบัญรูปภาพ	3
สารบัญตาราง	5
บทสรุปผู้บริหาร	6
บทที่ 1 บทนำ	9
1.1. ความสำคัญ / ที่มาของปัญหา	9
1.2. วัตถุประสงค์ของโครงการ	9
1.3. ขอบเขตของโครงการ	9
1.4. ประโยชน์ที่ได้รับของโครงการ	10
บทที่ 2 ทบทวนวรรณกรรม	11
2.1. ความรู้พื้นฐาน	11
2.2. งานที่เกี่ยวข้อง	13
2.3. ตารางเปรียบเทียบโครงการกับงานอื่นที่มีอยู่ในปัจจุบัน	16
บทที่ 3 การดำเนินงานโครงการ	17
3.1. การรวบรวมข้อมูล (Data Collection / Data Acquisition)	17
3.2. การเตรียมข้อมูลเบื้องต้น (Data Preprocessing / Data Cleaning)	25
3.3. การวิเคราะห์ภาพรวมข้อมูล (Exploratory Data Analysis / Data Visualization)	28
3.4. การสร้างแบบจำลอง (Model Building)	36
3.5. การนำแบบจำลองไปใช้งาน (Model Deployment)	43
บทที่ 4 ความเชื่อมโยงกับวิชาต่าง ๆ ในโมดูล	44
4.1 วช.310 การสำรวจและการเตรียมข้อมูล	44
4.2 วช.311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล	45
4.3 วช.312 ระบบธุรกิจอัจฉริยะ	50
4.4 วช.313 การวิเคราะห์การตลาด	53
บทที่ 5 บทสรุป	56
บรรณานุกรม	57

สารบัญรูปภาพ

รูปที่ 3-1 ภาพแสดง Data pipeline ของข้อมูลที่ได้จาก Web Scraping	17
รูปที่ 3-2 ภาพแสดงตัวอย่างหน้าเว็บไซต์ Thairath.com	19
รูปที่ 3-3 ภาพแสดงการ import library	20
รูปที่ 3-4 ภาพแสดงตัวอย่างการใช้เครื่องมือ Selenium (1)	20
รูปที่ 3-5 ภาพแสดงตัวอย่างการใช้เครื่องมือ Selenium (2)	21
รูปที่ 3-6 ภาพแสดงตัวอย่างการใช้เครื่องมือ BeautifulSoup	22
รูปที่ 3-7 ภาพแสดงตัวอย่างข้อมูลที่ได้จาก Web Scraping	22
รูปที่ 3-8 ภาพแสดงโค้ดการเปลี่ยน format ของเวลา	23
รูปที่ 3-9 ภาพแสดงโค้ดการเซฟเป็น CSV	23
รูปที่ 3-10 ภาพแสดงตัวอย่างข้อมูลหลังจาก Prepare data	23
รูปที่ 3-11 ภาพแสดง Data pipeline ของข้อมูลสำหรับทำ Model & Dashboard	23
รูปที่ 3-12 ภาพแสดง Data pipeline ของข้อมูลสำหรับการจำแนกเด็กยากจน/ไม่ยากจน	24
รูปที่ 3-13 ภาพแสดงตัวอย่างข้อมูลสาเหตุการออกจากระบบการศึกษาของนักเรียน	26
รูปที่ 3-14 ภาพแสดงข้อมูลสรุปสาเหตุหลักและสาเหตุการออก	27
รูปที่ 3-15 ภาพแสดงตัวอย่างโค้ดการเขียนเงื่อนไข	27
รูปที่ 3-16 ภาพแสดงภาพรวมของข้อมูลที่ใช้ในการทำโมเดล	28
รูปที่ 3-17 ภาพแสดงภาพรวมจำนวนเด็กยากจนพิเศษ	29
รูปที่ 3-18 ภาพแสดงภาพรวมจำนวนเด็กยากจนพิเศษในแต่ละภาค	29
รูปที่ 3-19 ภาพแสดงจังหวัดที่มีนักเรียนยากจนพิเศษมากที่สุด 10 อันดับแรก	30
รูปที่ 3-20 ภาพแสดงจำนวนเด็กที่หลุดออกจากระบบการศึกษา	31
รูปที่ 3-21 ภาพแสดงภาพรวมจำนวนเด็กหลุดออกจากระบบการศึกษาในแต่ละภาค	31
รูปที่ 3-22 ภาพแสดงจังหวัดที่มีเด็กหลุดออกจากระบบการศึกษา 10 อันดับ	32
รูปที่ 3-23 ภาพแสดงรายได้เฉลี่ยต่อครัวเรือนในแต่ละภาค	32
รูปที่ 3-24 ภาพแสดงรายได้เฉลี่ยต่อครัวเรือนในแต่ละจังหวัด	33
รูปที่ 3-25 ภาพแสดงหนี้สินเฉลี่ยต่อครัวเรือนในแต่ละภาค	33
รูปที่ 3-26 ภาพแสดงหนี้สินเฉลี่ยต่อครัวเรือนในแต่ละจังหวัด	34
รูปที่ 3-27 ภาพแสดงหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือนในแต่ละภาค	34
รูปที่ 3-28 ภาพแสดงหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือนในแต่ละจังหวัด	35
รูปที่ 3-29 ภาพแสดงงบประมาณด้านศึกษาในแต่ละภาค	35
รูปที่ 3-30 ภาพแสดงงบประมาณด้านศึกษาในแต่ละจังหวัด	36

รูปที่ 3-31 ภาพโลโก้ Pycaret และ Features ที่สำคัญ	38
รูปที่ 3-32 Machine Learning library สำหรับการทำ Regression ที่ Pycaret รองรับ	38
รูปที่ 3-33 ภาพสูตรการคำนวณ Robust Scaler	39
รูปที่ 3-34 ภาพอธิบายหลักการ One-Hot Encoding	39
รูปที่ 3-35 ภาพแสดงข้อมูลทั้งหมดหลังการทำ Data preprocessing	39
รูปที่ 3-36 ภาพแสดงข้อมูลรายละเอียดการ Setup() Model	40
รูปที่ 3-37 ภาพแสดงผลลัพธ์การประเมิน Model 1 ในแต่ละเทคนิค	41
รูปที่ 3-38 ภาพแสดงผลลัพธ์การประเมิน Model 2 ในแต่ละเทคนิค	42
รูปที่ 3-39 ภาพสูตรการคำนวณ MAE	43
รูปที่ 3-40 ภาพสูตรการคำนวณ MAPE	43
รูปที่ 3-41 ภาพสูตรการคำนวณ R2	43
รูปที่ 4-1 ภาพตัวอย่างข้อมูลที่ได้จากการ Web Scraping	45
รูปที่ 4-2 ภาพแสดงตัวอย่างของการแสดงผลข้อมูลในรูปแบบ Visualization โดยการทำ WordCloud	45
รูปที่ 4-3 ภาพแสดงจำนวนของสภาวะความยากจนของนักเรียน	46
รูปที่ 4-4 ภาพแสดงจำนวนข้อมูลสรุปผลรวมของสภาวะความยากจน จำแนกตามความคิดเห็นเพิ่มเติม	46
รูปที่ 4-5 Correlation Matrix แสดงความสัมพันธ์ของตัวแปรทั้งหมด	47
รูปที่ 4-6 ภาพแสดงผลลัพธ์ความแม่นยำของ Model 1	47
รูปที่ 4-7 ภาพแสดง Feature Important plot ของ Model 1	48
รูปที่ 4-8 ภาพแสดง SHAP Value ของ Model 1	48
รูปที่ 4-9 ภาพแสดงผลลัพธ์ความแม่นยำของ Model 2	49
รูปที่ 4-10 ภาพแสดง Feature Important plot ของ Model 2	49
รูปที่ 4-11 ภาพแสดง SHAP Value ของ Model 2	50
รูปที่ 4-12 ภาพ Dashboard แสดงภาพรวมของเด็กหลุดออกจากระบบการศึกษา	51
รูปที่ 4-13 ภาพแสดงตารางจำนวนเด็กยากจนพิเศษในแต่ละภูมิภาค	51
รูปที่ 4-14 ภาพแสดงตารางจำนวนเด็กยากจนพิเศษในแต่ละจังหวัดของภาคตะวันออกเฉียงเหนือ	51
รูปที่ 4-15 ภาพ Dashboard แสดงจำนวนนักเรียนยากจนและนักเรียนยากจนพิเศษ	52
รูปที่ 4-16 ภาพ Dashboard แสดงภาพรวมข้อมูลเปรียบเทียบรายได้และหนี้สินโดยเฉลี่ยในระดับภูมิภาค	53
รูปที่ 4-17 ภาพแผนที่แสดงความหนาแน่นของจำนวนเด็กยากจนพิเศษ	54
รูปที่ 4-18 ภาพแสดงจำนวนเด็กยากจนและยากจนพิเศษ 10 อันดับแรก	54

สารบัญตาราง

ตารางที่ 2-1 ตารางเปรียบเทียบโครงงานกับงานอื่น	16
ตารางที่ 3-1 แสดงหัวข้อตารางโดยมีหัวข้อข่าว แหล่งที่มา และ เทคนิคที่ใช้	18
ตารางที่ 3-2 แสดงข้อมูล ประเภทของไฟล์ วัตถุประสงค์ และแหล่งที่มาของข้อมูลทุติยภูมิ	25
ตารางที่ 3-3 รายละเอียดของชุดข้อมูลที่จะใช้แต่ละ Model	37
ตารางที่ 4-1 ผลลัพธ์ต่าง ๆ ที่ได้จากการทำ Web Scraping	44
ตารางที่ 4-2 ตารางแสดงจังหวัดที่ได้รับการพิจารณาการให้ทุน 10 อันดับแรก	55

บทสรุปผู้บริหาร

ความสำคัญ

การศึกษาเป็นปัจจัยสำคัญสำหรับการพัฒนาประเทศ เนื่องจากเป็นกลไกหนึ่งในการพัฒนาทรัพยากรมนุษย์ให้มีคุณภาพและสามารถสร้างความก้าวหน้าในการดำรงชีวิตต่อไปได้ หากเด็กไทยได้รับการศึกษาที่ดีก็จะเป็นหนึ่งในทรัพยากรที่สามารถพัฒนาประเทศไปในทิศทางที่ดีขึ้นได้เช่นเดียวกัน ซึ่งประเทศไทยในขณะนี้กำลังประสบปัญหาเด็กหลุดออกจากระบบการศึกษาที่ถึงแม้จะมีแนวทางการแก้ไขและช่วยเหลืออยู่บ้างแล้ว แต่ก็ยังมีเด็กบางส่วนที่ยังไม่กลับเข้าระบบการศึกษาอยู่ ปัญหาดังกล่าวจะส่งผลกระทบต่ออนาคตของเด็กโดยตรง และยังส่งผลกระทบต่อการพัฒนาประเทศในระยะยาวอีกด้วย ดังนั้น ปัญหาเด็กหลุดจากระบบการศึกษาจึงเป็นปัญหาสำคัญที่สมควรจะต้องได้รับการแก้ไขและป้องกันไม่ให้เกิดขึ้นอย่างจริงจัง เพื่อให้เด็กซึ่งเป็นทรัพยากรสำคัญของประเทศได้เติบโตไปอย่างมีคุณภาพชีวิตที่ดี พร้อมทั้งจะพัฒนาประเทศต่อไป

วัตถุประสงค์

เพื่อศึกษาปัญหาและปัจจัยที่ทำให้เด็กหลุดจากระบบการศึกษา พร้อมทั้งหาแนวทางการป้องกันไม่ให้นักเรียนที่กำลังศึกษาเล่าเรียนอยู่หลุดออกจากระบบการศึกษาในอนาคตและให้การช่วยเหลือเด็กที่หลุดจากระบบการศึกษาไปแล้วให้กลับเข้าระบบเพื่อศึกษาต่อในระดับต่อไป

Web Scraping

โดยได้ทำการรวบรวมข้อมูลจากแหล่งข่าวที่เกี่ยวข้องกับปัญหาของเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์ต่าง ๆ โดยใช้เทคนิค Web Scraping ออกมาเป็นไฟล์ Excel แล้วนำข้อมูลนั้นมาสร้าง WordCloud เพื่อดูคำที่ปรากฏในข้อมูลทั้งหมด โดยจะเรียงจากคำที่มีมากที่สุดไปจนถึงคำที่มีน้อยที่สุดและเชื่อมโยงว่ามีคำหรือปัจจัยใดที่เกี่ยวข้องกับปัญหาเด็กหลุดออกจากระบบการศึกษามากที่สุด โดยผลลัพธ์ของคำที่ออกมาสามารถบอกได้ว่า “ความยากจน” มีผลกระทบกับปัญหานี้มากที่สุด

การจำแนกเด็กยากจน/ไม่ยากจน

จากชุดข้อมูลเด็กหลุดออกจากระบบการศึกษาที่ได้มาจากรฐานข้อมูลเพื่อความเสมอภาคทางการศึกษา และโครงการพาน้องกลับมาเรียน จำนวนทั้งสิ้น 5,298 records สามารถนำข้อมูลมาวิเคราะห์เพื่อจำแนกเด็กยากจนและคาดว่าไม่ยากจนได้จากการพิจารณา Keyword ในคอลัมน์สาเหตุการออก ความต้องการ และความคิดเห็นเพิ่มเติม ผ่านการใช้เทคนิคเขียนเงื่อนไขโดยใช้ Keywords เพื่อจำแนกเด็กที่คาดว่ายากจน

จากการใช้เทคนิคดังกล่าวสามารถสรุปผลได้ว่า นักเรียนที่หลุดออกจากระบบการศึกษาส่วนใหญ่เป็นเด็กที่มีสถานะยากจนและยังนำมาสู่สมมติฐานที่ว่า “ความยากจน” เป็นปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุด

Regression Model

จากสมมติฐานที่ว่าปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุด คือ ความยากจน เมื่อทำการรวบรวมชุดข้อมูลต่าง ๆ ที่เกี่ยวข้องกับปัจจัยความยากจนออกมาในรูปแบบไฟล์ Excel จะประกอบด้วยข้อมูลจาก 3 แหล่ง คือ สำนักงานสถิติแห่งชาติ, ฐานข้อมูลเพื่อความเสมอภาคทางการศึกษา และโครงการพาน้องกลับมาเรียน โดยชุดข้อมูลที่ได้จะมีทั้งหมด 4 ชุดข้อมูล ได้แก่ ข้อมูลเด็กหลุดออกจากระบบการศึกษา เดือนกรกฎาคม ปี 2565, ข้อมูลจำนวนนักเรียนยากจนและยากจนพิเศษ ปี 2564, รายได้เฉลี่ยต่อเดือนของครัวเรือน รายจังหวัด ปี 2564 และหนี้สินเฉลี่ยต่อครัวเรือน จำแนกตามวัตถุประสงค์ของการกู้ยืม รายจังหวัด ปี 2564

เมื่อทำการรวมชุดข้อมูลทั้งหมดเป็นชุดข้อมูลเดียวกันเพื่อนำไปสร้าง Model ทำนายจำนวนเด็กหลุดออกจากระบบการศึกษา โดยเลือกใช้ library Pycaret ฟังก์ชัน Compare_Model() ในการเลือกเทคนิคของ Model ที่มีผลลัพธ์ที่ดีที่สุด จึงได้ข้อสรุปว่า Model ที่ใช้ คือ Extra Trees regressor และให้ค่าความแม่นยำดังนี้

$$R2 = 0.8065, MAE = 4742.2 \text{ และ } MAPE 0.7658$$

โดยสิ่งที่ได้จาก Model คือ หากจำนวนนักเรียนยากจนพิเศษเพิ่มขึ้นจะทำให้จำนวนนักเรียนหลุดออกจากระบบการศึกษามากขึ้นตามไปด้วย

จึงนำมาสู่ Model ที่ 2 Model ทำนายจำนวนนักเรียนยากจนพิเศษ เช่นเดียวกับ Model ที่ 1 ใช้ฟังก์ชัน Compare_Model() ของ Pycaret ในการเลือกเทคนิคสำหรับสร้าง Model โดยเทคนิคที่ใช้ คือ Random Forest และมีค่าความแม่นยำ ดังนี้

$$R2 = 0.8065, MAE = 4742.2 \text{ และ } MAPE 0.7658$$

โดย Model ทำให้ทราบว่า หากเป็นจังหวัดที่อยู่ในภาคตะวันออกเฉียงเหนือจะมีความโอกาสที่จะมีนักเรียนยากจนพิเศษเป็นจำนวนมาก และหากมีรายได้เฉลี่ยต่อครัวเรือนต่ำก็จะทำให้จำนวนนักเรียนยากจนพิเศษมีจำนวนมากอีกด้วย

สรุปผลการวิเคราะห์

จากการวิเคราะห์ทั้งหมดสามารถสรุปได้ว่าความยากจนเป็นปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุด หากต้องการลดจำนวนเด็กหลุดออกจากระบบการศึกษา ก็จำเป็นที่จะต้องทำการลดจำนวนเด็กยากจนพิเศษก่อน ดังนั้น แนวทางการลดจำนวนเด็กหลุดออกจากระบบที่คณะผู้จัดทำเสนอ คือ การให้ทุนการศึกษา โดยพิจารณาการให้ทุนการศึกษาจากจังหวัดที่มีจำนวนเด็กยากจนพิเศษมากที่สุด คือ จังหวัดอุบลราชธานี ก่อน

ประโยชน์ที่กลุ่มเป้าหมายและส่วนรวมจะได้รับ

จากการทำโครงการ สามารถนำข้อมูลที่ได้ไปใช้ประโยชน์ในการแก้ไขและป้องกันปัญหาเด็กหลุดจากระบบการศึกษาได้ ซึ่งจะส่งผลให้จำนวนเด็กหลุดจากระบบการศึกษาลดลง พร้อมทั้งเพิ่มโอกาสที่เด็กจะได้รับการศึกษาขั้นพื้นฐานอย่างทั่วถึง โดยการนำข้อมูลไปพิจารณาการให้ทุนการศึกษากับเด็กที่มีปัญหาดังกล่าวจะสามารถลดปัญหาเรื่องความเหลื่อมล้ำภายในประเทศไทยได้ และเด็กทุกคนจะสามารถเติบโตไปได้อย่างมีคุณภาพชีวิตและอนาคตที่ดี พร้อมทั้งจะเป็นกำลังสำคัญในการพัฒนาประเทศต่อไป

บทที่ 1

บทนำ

1.1. ความสำคัญ / ที่มาของปัญหา

เด็กและเยาวชนเป็นทรัพยากรที่มีคุณค่า อีกทั้งยังมีความสำคัญอย่างมากกับการพัฒนาประเทศให้เดินหน้าต่อไป จึงจำเป็นที่จะต้องดูแลเอาใจใส่อย่างรอบด้าน ทั้งจากครอบครัว สถานศึกษา และสังคม ซึ่งการศึกษาถือเป็นส่วนสำคัญอีกส่วนหนึ่งในการวางรากฐานและพัฒนาให้เด็กสามารถก้าวไปสู่การเป็นประชากรที่มีคุณภาพ เมื่อเด็กเข้ารับการศึกษาแล้วจะต้องพยายามให้เด็กเล่าเรียนจนจบการศึกษาขั้นพื้นฐานและได้ศึกษาในระดับที่สูงขึ้นต่อไป แต่ปัญหาที่เกิดขึ้นมาโดยตลอด คือ ปัญหาเด็กหลุดจากระบบการศึกษา ในแต่ละปีประเทศไทยมีเด็กหลุดจากระบบการศึกษามากถึง 2% จากจำนวนนักเรียนทั้งหมด ซึ่งถือว่าเป็นจำนวนที่มากพอสมควร ปัญหาเด็กหลุดจากระบบการศึกษาจะส่งผลกระทบโดยตรงต่อทางเลือกในการดำเนินชีวิต การงาน และอนาคต จึงเป็นปัญหาที่มีความสำคัญและเป็นปัญหาอันดับต้น ๆ ที่ควรได้รับการแก้ไขอย่างเร่งด่วนเพื่อป้องกันไม่ให้เกิดปัญหาเด็กหลุดจากระบบการศึกษาเพิ่มขึ้นอีก

ทางคณะผู้จัดทำจึงเล็งเห็นความสำคัญในการปรับปรุงแก้ไขปัญหาเด็กหลุดจากระบบการศึกษาและเพื่อเป็นการยกระดับขีดความสามารถทางการบริหารจัดการด้านการศึกษา จึงมีความสนใจที่จะศึกษาปัญหาและปัจจัยที่ทำให้เด็กหลุดจากระบบการศึกษา เพื่อหาแนวทางการช่วยเหลือและป้องกันไม่ให้นักเรียนคนไหนหลุดจากระบบการศึกษาอีก และหวังเป็นอย่างยิ่งว่าแนวโน้มที่เด็กหลุดจากระบบการศึกษาจะลดลงในระยะยาว

1.2. วัตถุประสงค์ของโครงการ

1. เพื่อศึกษาปัญหาและปัจจัยที่ทำให้เด็กหลุดจากระบบการศึกษา
2. เพื่อหาแนวทางการช่วยเหลือและป้องกันไม่ให้นักเรียนที่กำลังศึกษาเล่าเรียนอยู่หลุดออกจากระบบการศึกษาในอนาคต พร้อมทั้งให้การช่วยเหลือเด็กที่หลุดออกจากระบบการศึกษาไปแล้วให้กลับเข้าสู่ระบบการศึกษา

1.3. ขอบเขตของโครงการ

ขอบเขตด้านประชากร

ในการวิจัยนี้ ได้ทำการศึกษากลุ่มประชากรเป้าหมายซึ่งเป็นนักเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.)

ขอบเขตด้านเนื้อหา

ในการวิจัยนี้ ได้ทำการศึกษาค้นคว้าแนวคิด ทฤษฎีและผลงานวิจัยจากสื่อออนไลน์ภายในประเทศที่เกี่ยวข้องกับปัญหาเด็กหลุดจากระบบการศึกษาและการแก้ปัญหาดังกล่าว

1.4. ประโยชน์ที่ได้รับของโครงการ

มีความรู้ความเข้าใจเกี่ยวกับปัญหาและปัจจัยที่ทำให้เด็กหลุดจากระบบการศึกษามากยิ่งขึ้น สามารถนำความรู้ที่ได้ไปต่อยอดในการพัฒนาแนวทางเพื่อแก้ไขปัญหาคือหลุดจากระบบการศึกษาให้มีประสิทธิภาพมากขึ้นในอนาคต และเพื่อป้องกัน สร้างความมั่นใจว่าเด็กทุกคนจะสามารถเข้ารับการศึกษาขั้นพื้นฐานได้อย่างเท่าเทียม

บทที่ 2

บททวนวรรณกรรม

2.1. ความรู้พื้นฐาน

สถานการณ์ความยากจน

สถานการณ์ความยากจนในปี 2564 ปรับตัวดีขึ้น คนยากจนมีจำนวนทั้งสิ้น 4.4 ล้านคน คิดเป็นสัดส่วนคนยากจนที่ร้อยละ 6.32 ซึ่งลดลงจากปีก่อนที่มีสัดส่วนคนยากจนร้อยละ 6.83 ในขณะที่เมื่อพิจารณาจำนวนครัวเรือนยากจน พบว่าในปี 2564 ครัวเรือนยากจนมีจำนวนทั้งสิ้น 1.24 ล้านครัวเรือน คิดเป็น สัดส่วนร้อยละ 4.79 ของครัวเรือนทั้งหมด ลดลงเมื่อเทียบกับปี 2563 ที่มีจำนวนครัวเรือนยากจน ประมาณ 1.40 ล้านครัวเรือน สำหรับสาเหตุสำคัญที่สถานการณ์ความยากจนปรับตัวดีขึ้นเป็นผลมาจาก ในช่วงปี 2563 - 2564 ที่รัฐบาลมีการดำเนินมาตรการเพื่อช่วยเหลือและบรรเทาผลกระทบที่เกิดขึ้นจากปัญหาวิกฤตการณ์การแพร่ระบาดของ COVID-19 อย่างต่อเนื่องโดยเฉพาะกับกลุ่มคนยากจนและผู้มีรายได้น้อย

เมื่อพิจารณาในด้านมิติพื้นที่และเศรษฐกิจของคนจนพบว่า ภาคใต้มีปัญหาความยากจนรุนแรงที่สุด โดยภาคใต้มีสัดส่วนคนจนสูงสุดที่ร้อยละ 10.9 ขณะที่ภาคตะวันออกเฉียงเหนือมีจำนวนคนจนมากที่สุด โดยเป็นภูมิภาคที่มีรายได้อ่อนแอสุดมาโดยตลอด ซึ่งจังหวัดที่มีรายได้อ่อนแอที่สุดของประชากรต่ำมีโอกาสที่จะมีปัญหาความยากจนมากกว่าจังหวัดที่มีรายได้อ่อนแอสูง สะท้อนถึงการพัฒนาของพื้นที่ที่มีผลสำคัญต่อการแก้ไขปัญหาความยากจน คนจนมีอัตราการพึ่งพิง (จำนวนเด็กอายุต่ำกว่า 15 ปี และผู้สูงอายุต่อวัยแรงงาน) สูงกว่าครัวเรือนไม่ยากจนอย่างชัดเจนและครัวเรือนขนาดใหญ่จะมีปัญหาความยากจนมากกว่าครัวเรือนขนาดเล็ก นอกจากนี้ ยังพบว่าในอดีตครัวเรือนเกษตรมักเป็นครัวเรือนยากจน แต่ปัจจุบันครัวเรือนยากจน ส่วนใหญ่เป็นครัวเรือนไม่มีส่วนร่วมทางเศรษฐกิจ คิดเป็นสัดส่วนร้อยละ 33.99 ของครัวเรือนยากจนทั้งหมด ส่วนหนึ่งเป็นผลมาจากการเปลี่ยนแปลงโครงสร้างประชากรสู่สังคมสูงวัย ซึ่งพบว่าผู้สูงวัยไม่สามารถหารายได้ และเข้าถึงหลักประกันรายได้หลังเกษียณ

สถานการณ์ความเหลื่อมล้ำ

ความเหลื่อมล้ำด้านรายได้ปรับตัวเพิ่มขึ้น โดยค่าสัมประสิทธิ์จีนีอยู่ที่ 0.430 ปรับตัวเพิ่มขึ้นจาก 0.429 ในปี 2562 กลุ่มคนที่มีเศรษฐกิจดีที่สุด (decile 10) มีส่วนแบ่งรายได้เพิ่มสูงขึ้นจากร้อยละ 33.17 ของรายได้รวมทั้งประเทศ ในปี 2562 เป็นร้อยละ 33.43 หรือคิดเป็นส่วนแบ่งรายได้สูงกว่ากลุ่มที่มีเศรษฐกิจต่ำที่สุด (decile 1) ถึง 16.4 เท่า เมื่อพิจารณาโครงสร้างแหล่งที่มาของรายได้ที่เป็นตัวเงินในแต่ละกลุ่มรายได้พบว่า โครงสร้างรายได้ของประชากรกลุ่มที่มีเศรษฐกิจดีที่สุด (decile 10) และกลุ่มที่มีเศรษฐกิจต่ำที่สุด (decile 1) แตกต่างกันอย่างชัดเจน โดยครึ่งหนึ่งของรายได้ในกลุ่มที่มีเศรษฐกิจต่ำที่สุด (decile 1) มาจากเงินโอน ซึ่งแบ่งเป็นเงินช่วยเหลือจากภาครัฐร้อยละ 30.0 และเงิน

ช่วยเหลือจากบุคคลอื่นร้อยละ 21.0 ขณะที่รายได้ จากการทำงานอยู่ที่ร้อยละ 48.5 ตรงข้ามกับกลุ่มคนที่มีเศรษฐานะดีที่สุด (decile 10) ที่มีรายได้หลักมาจาก การทำงานอยู่ที่ร้อยละ 83.1 สะท้อนให้เห็นถึงความไม่แน่นอนของรายได้ในกลุ่มประชากรที่มีฐานะทางเศรษฐกิจต่ำที่พึ่งพิงรายได้จากเงินโอนในระดับสูงแตกต่างจากกลุ่มประชากรที่มีฐานะทางเศรษฐกิจสูงที่มีรายได้หลักมาจากการทำงาน

ภาพรวมความเหลื่อมล้ำด้านโอกาสทางการศึกษา

ความเหลื่อมล้ำด้านโอกาสทางการศึกษามีทิศทางที่ดีขึ้นในเกือบทุกระดับชั้น กลุ่มเด็กที่มีเศรษฐานะดีที่สุด (decile 10) มีอัตราการเข้าศึกษาต่อระดับมัธยมศึกษาตอนปลายร้อยละ 80.9 สูงกว่ากลุ่มเด็กที่มีเศรษฐานะต่ำที่สุด (decile 1) ประมาณ 1.6 เท่า แม้ว่าจะมีทิศทางที่ดีขึ้นต่อเนื่องจากสองปีก่อนหน้าแต่สถานการณ์ การแพร่ระบาดของ COVID-19 ทำให้มีนักเรียนตกหล่น/ออกกลางคันจำนวน 28,793 คน โดยอยู่ในระดับอนุบาล-ประถมศึกษาร้อยละ 49.0 และอยู่ในระดับมัธยมศึกษาร้อยละ 51.1 นอกจากนี้ยังส่งผลให้เกิดปัญหาภาวะสูญเสียการเรียนรู้ (Learning Loss) โดยเฉพาะในกลุ่มเด็กเล็ก ความเหลื่อมล้ำด้านคุณภาพและการจัดสรรทรัพยากรทางการศึกษายังคงมีอยู่ในหลายมิติ คุณภาพการศึกษามีความเชื่อมโยงกับขนาดโรงเรียน ที่ตั้งและภูมิภาค โดยโรงเรียนขนาดใหญ่มีคะแนนผลการทดสอบทางการศึกษาสูงกว่าโรงเรียนขนาดเล็กและโรงเรียนในพื้นที่กรุงเทพมหานครมีคะแนนผลการทดสอบสูงกว่าภูมิภาคอื่น สาเหตุส่วนหนึ่งเกิดจากนโยบายของรัฐทั้งด้านงบประมาณและบุคลากร จัดสรรตามขนาดโรงเรียน จึงเอื้อต่อโรงเรียนขนาดใหญ่มากกว่าโรงเรียนขนาดเล็ก

ปัจจัยที่ทำให้เด็กหลุดออกจากการศึกษา

กฎระเบียบ/หลักสูตรไม่ตอบโจทย์

โดยทุกฝ่ายต้องมีเป้าหมายร่วมกันว่า ถ้าพบตัวนักเรียนแล้วตัวเด็กมีความพร้อม จะต้องช่วยพากลับโรงเรียนทันที ดังนั้นทุกโรงเรียนต้องรับรู้ร่วมกันว่า ปรากฏการณ์เชิงแรงของกฎระเบียบ เช่น การไม่สามารถรับเด็กเข้ากลางเทอมได้ คือ อุปสรรคสำคัญที่ทำให้เด็กหลุดมือไป เพราะแม้จะเป็นเวลาสั้น ๆ เพียงไม่กี่สัปดาห์หรือไม่กี่เดือนของการรอรอบปฏิทินการศึกษา แต่สำหรับเด็กที่พ้นจากรั้วโรงเรียนไปแล้ววิถีชีวิตทุกวันที่ผ่านไปของเขาจะยิ่งเบนออกไปจากการศึกษา ทำให้รู้สึกหมดกำลังใจกับการเรียน

ขาดโรงเรียนที่พอเพียง/คุณภาพครูที่เพียงพอ

ในพื้นที่ทุรกันดารห่างไกล ปัญหาการคมนาคม คือ อุปสรรคที่ไม่ได้เพียงลดทอนกำลังใจของเด็กเท่านั้น แต่ในเชิงรูปธรรมแล้ว มันได้ตัดขาดเด็ก ๆ ออกจากโรงเรียนในหลายช่วงเวลาของปี เนื่องจากบนดอยสูง ในป่าเขา หรือเกาะห่างไกล พวกเขาต้องเดินเท้าเป็นระยะทางไกล ต่อสู้กับภัยธรรมชาติทั้งน้ำหลาก ดินถล่ม นานากรรม หรือกระทั่งความเสี่ยงจากสัตว์ร้าย การขาดแคลนโรงเรียนคุณภาพที่กระจายอยู่ไม่พอเพียงในหลายพื้นที่ จึงเป็นปัจจัยต้น ๆ ที่ทำให้เด็กหลายคนหลุดออกไป หรือส่วนหนึ่งที่ฝ่าฟันจนจบ ป.6 หรือ ม.3 ได้ ก็หมดหนทางไปต่อในช่วงชั้นที่สูงขึ้น เพราะไม่มีโรงเรียนใกล้บ้านรองรับ

ขาดที่ปรึกษาช่วยชี้ทาง/ผลลัพธ์ทางการศึกษาไม่สร้างแรงจูงใจ

แม้ว่าภูมิหลังของเด็กที่ออกนอกระบบส่วนใหญ่จะมาจากความยากจน แต่หากเราส่องลึกลงไปภายในแกนของปัญหาจะเห็นว่าความเปราะบางในใจเด็กถือว่ามีส่วนอย่างมากที่จะช่วยรั้งเขาไว้ในระบบหรือผลักดันให้หลุดออกไป การรับเด็กเข้ามาสู่ระบบการศึกษา ไม่ว่าจะเป็นในหรือนอกโรงเรียน จึงต้องมี Case Manager ที่ดูแลเด็กได้เป็นรายคน และเป็นที่พึ่งให้กับครอบครัวของเขาได้ด้วย การมีที่ปรึกษาใกล้ชิดจะทำให้เด็กอุ่นใจเชื่อมั่นว่ามีคนที่เข้าใจและให้คำแนะนำได้กับการตัดสินใจในทุกทางแยกสุมเสี่ยงของชีวิต โดยการทำงานกับเด็กบางกลุ่มที่ปรึกษาจำเป็นต้องมีความสามารถด้านภาษาถิ่น เข้าใจบริบทวัฒนธรรมความเชื่อต่าง ๆ เพื่อการทำงานที่ลงลึกในเชิงคุณภาพ พร้อมเข้าไปประคองเด็กได้ในทุกช่วงเวลาเพื่อไม่ให้เขาหลุดออกไป

ตะเข็บรอยต่อโรงเรียนต่างพื้นที่/ขาดความพร้อมในการเรียนช่วงโรงเรียนปิดเพราะสถานการณ์โรคระบาด

จากสถานการณ์ COVID-19 ทำให้มีครอบครัวที่สูญเสียงาน ต้องอพยพย้ายถิ่นไปทำงานในพื้นที่อื่นหรือตัดสินใจกลับภูมิลำเนา หน่วยงานที่เกี่ยวข้องจึงจำเป็นต้องคำนึงถึงโจทย์สำคัญว่าเมื่อเด็กและเยาวชนต้องย้ายถิ่นตามผู้ปกครองไปเด็กจำนวนมากต้องออกจากโรงเรียนกลางคัน ตัวเด็กจะเข้าศึกษาต่อได้อย่างไร เพราะบางคนอาจมีเหตุผลที่ต้องย้ายไปในช่วงกลางปีการศึกษา หรือจบชั้น ป.6 หรือ ม.3 แล้วตัวเด็กขาดความพร้อมในการเรียนต่อ ตรงนี้จำเป็นต้องมีระบบช่วยเหลือและประชาสัมพันธ์ข้อมูลให้ไปถึงผู้ปกครอง เพื่อไม่ให้เด็กต้องหลุดไปเพราะเข้าไม่ถึงความรู้ในขั้นตอนการพาลูกหลานเข้าโรงเรียน หรือโรงเรียนปลายทางไม่มีระบบรองรับนักเรียนกลางเทอม

ขาดความพร้อมในการเรียนออนไลน์ ไม่มีเครื่องมืออุปกรณ์ พื้นที่และผู้อำนวยความสะดวกที่เหมาะสมกับสถานการณ์

จะส่งผลให้เด็กมีโอกาสหลุดจากการศึกษาเพิ่มมากขึ้น หรือแม้แต่คนที่ตามกลับเข้าสู่ระบบการศึกษาได้แล้วก็ยังมีความเสี่ยงที่จะหลุดออกจากระบบเช่นเดิม อีกประเด็นซึ่งเรื่องเหล่านี้เราไม่สามารถโยนภาระการดูแลแก้ไขให้หน่วยงานใดหนึ่งได้ แต่ต้องเชื่อมรอย-ประสานชุมชน ท้องถิ่น ภาครัฐ เอกชน ภาคประชาสังคม เพื่อร่วมกันหาแนวทางแก้ไขที่ชัดเจน เพราะคนที่รู้สถานการณ์หรือตัวเลขต่าง ๆ คือหน่วยงานที่ทำงานในพื้นที่ ดังนั้น ถ้ามีการส่งต่อข้อมูลเพื่อออกแบบการทำงานตามความถนัดของแต่ละหน่วยได้ก็จะสามารถสร้างแพลตฟอร์มที่ส่งต่อถึงกันได้ทั้งหมดและถ้ามีเด็กออกจากโรงเรียนกลางคัน ยกตัวอย่างเช่น จังหวัดกรุงเทพฯ เมื่อตัวเด็กไปอยู่ที่อื่นก็จะยังเข้าเรียนต่อได้โดยไม่ต้องหลุดจากการศึกษาไปเสียเลย ๆ

2.2. งานที่เกี่ยวข้อง

โครงการ “พาน้องกลับมาเรียน”

โครงการนี้ดำเนินการโดย 3 หน่วยงานหลักของกระทรวงศึกษาธิการ (ศธ.) ได้แก่ สำนักงาน ปลัดกระทรวงศึกษาธิการ (สป.) สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน(สพฐ.) และสำนักงาน คณะกรรมการการอาชีวศึกษา (สอศ.) พร้อมด้วย 11 พันมิตร ประกอบด้วยกระทรวงมหาดไทย กระทรวงสาธารณสุข กระทรวงวัฒนธรรม กระทรวงการท่องเที่ยวและกีฬา กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ สำนักงานตำรวจ แห่งชาติ สำนักงานพระพุทธศาสนาแห่งชาติ กองอำนวยการรักษาความมั่นคงภายในราชอาณาจักร กรุงเทพมหานคร และกองทุนเพื่อความเสมอภาคทางการศึกษา

เนื่องจากในช่วงเดือนพฤศจิกายนปี 2564 ที่ผ่านมา ประเทศไทยได้สำรวจพบว่า **มีเด็กตกหล่นที่ หลุดออกจากระบบการศึกษามากถึง 238,707 คน** รัฐบาลและทุกหน่วยงานพยายามหาแนวทาง ช่วยกันทุกวิถีทาง ที่จะสร้างโอกาสทางการศึกษาให้กับเด็กเหล่านี้ เพราะหลายคนที่ออกจากระบบ การศึกษา ไม่ใช่เฉพาะปัญหาจากการแพร่ระบาด COVID-19 แต่มีเหตุผลอื่น ๆ เช่น ความจำเป็นของ ครอบครัว เพื่อน หรือสิ่งแวดล้อมต่าง ๆ โดยจะมีการลงติดตามถึงบ้าน เพื่อตามเด็กเหล่านี้กลับสู่ระบบ การศึกษาอีกครั้ง

หลังจากได้เริ่มโครงการหลายหน่วยงานได้เริ่มต้นลงพื้นที่ปิกหมุดค้นหาเด็กหลุดออกนอกระบบ ไปบ้างแล้ว ทำให้ในช่วงเดือนมกราคม 2565 มีตัวเลขจำนวนนักเรียน นักศึกษา นักเรียนพิการ และผู้ พิการ ที่ตกหล่นและออกกลางคัน เหลือ **121,642 คน** ในจำนวนนี้เป็นนักเรียน นักศึกษา กลุ่มปกติ ในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) สำนักงานคณะกรรมการการอาชีวศึกษา (สอศ.) สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.) และสำนักงานส่งเสริมการศึกษานอก ระบบและการศึกษาตามอัธยาศัย (กศน.) **รวมทั้งสิ้น 67,129 คน พบตัวแล้ว 52,760 คน** ในจำนวนที่ พบตัวนี้ มีเด็กที่กลับเข้าระบบการศึกษา 31,446 คน ไม่กลับเข้าระบบ 21,314 คน อยู่ระหว่างการ ติดตาม 5,628 คน และติดตามแล้วไม่พบตัว 8,741 คน ส่วนที่เหลือเป็นกลุ่มนักเรียนพิการ สังกัดสำนัก บริหารงานการศึกษาพิเศษ (สศศ.) สพฐ.และส่วนใหญ่เป็นกลุ่มนักศึกษา กศน.อายุเกิน 18 ปี ที่เกินวัย การศึกษาภาคบังคับ และมีความต้องการประกอบอาชีพ

ปัจจัยสู่ความสำเร็จของการดำเนินงาน

เกิดจากการประสานการดำเนินงานร่วมกันระหว่างหน่วยงานต่าง ๆ ที่ลงพื้นที่และการแก้ไข ปัญหาอย่างจริงจัง มีการนำเทคโนโลยีเชื่อมโยงข้อมูล (Big Data) มาใช้เป็นเครื่องมือในการติดตาม ค้นหา และบันทึกผลการติดตาม ผ่านแอปพลิเคชัน “พาน้องกลับมาเรียน” และเว็บไซต์ dropout.edudev.in.th ที่สำคัญเกิดจากความร่วมแรงร่วมใจของผู้บริหาร ครู และบุคลากรการศึกษา ในการลงพื้นที่เยี่ยมบ้านเด็ก เพื่อค้นหาเด็กตกหล่นและเด็กออกกลางคัน ให้กลับเข้าสู่ระบบการศึกษา อย่างต่อเนื่อง

สำหรับปัญหาอุปสรรคที่พบมีหลายด้านเช่นกัน ไม่ว่าจะเป็น

ปัญหาด้านครอบครัว

เนื่องจากนักเรียนที่หลุดระบบการศึกษาส่วนใหญ่มีฐานะยากจน มีปัญหาด้านค่าใช้จ่าย ครอบครัวหย่าร้าง ต้องช่วยพ่อแม่ประกอบอาชีพหารายได้เลี้ยงครอบครัว ประกอบกับมีพี่น้องหลายคนอาศัยอยู่กับปู่ย่า ตายาย จึงทำให้ไม่สามารถเข้าเรียนในระบบได้

ปัญหาด้านการบันทึกข้อมูลไม่เป็นปัจจุบัน

นักเรียนได้ย้ายสถานศึกษาจากโรงเรียนเดิมไปแล้ว ซึ่งโรงเรียนได้ดำเนินการย้ายนักเรียนตามขั้นตอนดำเนินงานที่ถูกต้องแล้ว หรือนักเรียนจบการศึกษาในระดับชั้น ม.3 จากโรงเรียนขยายโอกาสฯ ไปแล้ว แต่ยังมีข้อมูลนักเรียนว่าเป็นนักเรียนออกกลางคัน

นักเรียนมีอายุเกินเกณฑ์การศึกษาภาคบังคับ

โรงเรียนได้จำหน่ายนักเรียนออกจากฐานข้อมูลนักเรียนไปแล้ว แต่ยังมีข้อมูลนักเรียนว่าเป็นนักเรียนออกกลางคัน

ข้อมูลนักเรียนไม่เชื่อมโยงกันอย่างเป็นระบบ

มีหลายฐานข้อมูล ส่งผลกระทบต่อการติดตามในเชิงพื้นที่

ขาดแคลนอุปกรณ์การเรียน

ในช่วง COVID-19 ทำให้เรียนไม่ทัน จึงหยุดเรียน ประกอบกับผู้ปกครองพาเด็กไปทำงานรับจ้าง ส่งผลกระทบต่อการติดตามตัวเด็ก

ความหลากหลายของชาติพันธุ์นักเรียน

กระเหรี่ยง มูเซอ แม้ว ไทยใหญ่ และประเทศเพื่อนบ้านที่อพยพมาจากพื้นที่ชายขอบที่มีวิถีทางขนบธรรมเนียมของแต่ละชนเผ่าที่เน้นให้ความสำคัญกับการดำรงชีวิต มากกว่าการมุ่งให้ความสำคัญในการศึกษา

ดังนั้น เพื่อให้การทำงานในเรื่องนี้เป็นไปตามเป้าหมาย กระทรวงศึกษาธิการจึงได้มีข้อเสนอที่จะทำให้อสำเร็จ ดังนี้

1. ต้องแนะนำให้นักเรียนเข้าเรียนการศึกษานอกระบบ (กศน.) หรือเรียนสายวิชาชีพ เพื่อจะได้ประกอบอาชีพหารายได้เลี้ยงครอบครัว และสามารถเรียนจนจบการศึกษาภาคบังคับหรือการศึกษาระดับพื้นฐาน
2. เร่งพัฒนาระบบการจัดเก็บข้อมูลนักเรียนในแอปพลิเคชัน Dropout ของ สพฐ. ให้เป็นระบบเดียวกัน เป็นปัจจุบัน และเชื่อมโยงกับข้อมูลของสถานศึกษาทุกสังกัด โดยใช้เลขประจำตัวประชาชน 13 หลัก เพื่อป้องกันการบันทึกข้อมูลซ้ำซ้อนของสถานศึกษา หรือการเปลี่ยนชื่อ-สกุลของนักเรียน
3. ให้นำหน่วยงานในระดับจังหวัดและภาคเข้ามาช่วยเรียกดูข้อมูลในภาพรวมของจังหวัดและภาคของตนเองในแอปพลิเคชัน Dropout เพื่อความรวดเร็วในการประสานงานและเก็บรวบรวมข้อมูล โดยมีการติดตามทุกภาคเรียน เพื่อความชัดเจนของข้อมูล

การทำนายระดับความยากจนจากของข้อมูลสำมะโนประชากรด้วยการเรียนรู้ของเครื่อง

PREDICTION OF POVERTY LEVEL ON CENSUS DATA USING MACHINE LEARNING

จุดมุ่งหมายของงานนี้ คือ พัฒนาแบบจำลองการทำนายระดับความยากจนของแต่ละบุคคล และหาปัจจัยที่ส่งผลกระทบต่อความยากจน ด้วยการนำเทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning) มาใช้ในการวิเคราะห์ข้อมูลสำมะโนประชากรเพื่อระบุระดับความยากจนในแต่ละบุคคล ซึ่งจะแบ่งระดับความยากจนเป็น 4 ระดับคือ บุคคลยากจนขั้นรุนแรง (Extreme poverty), บุคคลยากจนปานกลาง (Moderate poverty), บุคคลเสี่ยงจะยากจน (Vulnerable households) และบุคคลไม่เสี่ยงจะยากจน (Non vulnerable households)

โดยชุดข้อมูลที่ใช้ในงานนี้นำมาจากชุดข้อมูลประชากรประเทศอินโดนีเซียจำนวน 6,388 ครอบครัว ดังนั้น ผลลัพธ์ของงานวิจัยนี้เป็นการสะท้อนระดับความยากจนในประเทศอินโดนีเซียเท่านั้น โดยผู้ทำการวิจัยคิดว่าความยากจนเป็นแนวคิดหลายมิติ โดยที่ไม่สามารถพิจารณาเรื่องราวได้ว่าเป็นตัวทำนายความยากจนเพียงเรื่องเดียว แต่นำข้อมูลต่าง ๆ มาพิจารณาร่วมประกอบกัน เช่น ระดับการศึกษาสุขภาพร่างกาย และสภาพครัวเรือน เป็นต้น ซึ่งตัวแปรทั้งหมดเหล่านี้ ล้วนเป็นปัจจัยสำคัญที่สามารถพิจารณาได้ว่าบุคคลดังกล่าวมีระดับความยากจนในระดับใด

2.3. ตารางเปรียบเทียบโครงงานกับงานอื่นที่มีอยู่ในปัจจุบัน

ตารางที่ 2-1 ตารางเปรียบเทียบโครงงานกับงานอื่น

ลักษณะสำคัญ (Feature)	โครงการพำน้องกลับบ้าน	งานวิจัยการทำนายระดับความยากจน	งานของเรา
อัตโนมัติ	/	/	/
สามารถปรับแต่งโมเดลได้		/	/
ข้อมูลเฉพาะเจาะจง	/	/	/
นำ Machine Learning มาวิเคราะห์		/	/
มี Dashboard แสดงข้อมูลของสถานการณ์เด็กหลุดออกจากระบบการศึกษา	/		/
มีการพิสูจน์ว่าปัจจัยที่แท้จริงที่ทำให้เด็กออกจากระบบการศึกษา คือ ความยากจน			/
มีการเก็บข้อมูลจากข่าว โดยใช้เทคนิค Web Scraping			/

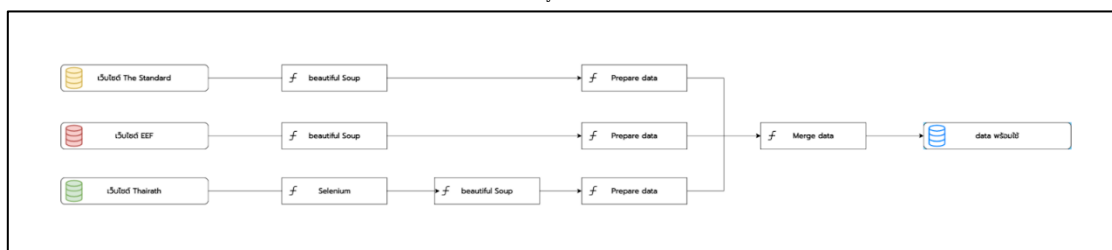
บทที่ 3

การดำเนินงานโครงการ

3.1. การรวบรวมข้อมูล (Data Collection/ Data Acquisition)

3.1.1 ข้อมูลปฐมภูมิ

รูปที่ 1



รูปที่ 3-1 ภาพแสดง Data pipeline ของข้อมูลที่ได้จาก Web Scraping

Web Scraping

ในส่วนของการรวบรวมข้อมูล (Data Collection) สามารถทำได้โดยการใช้เทคนิค Web Scraping ในการเก็บรวบรวมข้อมูลและแหล่งข่าวจากเว็บไซต์ต่าง ๆ แล้วนำไปใช้ในการสร้าง WordCloud เพื่อหาปัจจัยสำคัญที่ส่งผลให้เด็กหลุดออกจากระบบการศึกษา ในส่วนของ library ที่เลือกใช้ในการ Web Scraping มีอยู่ 2 library ดังนี้

1. Selenium คือ เครื่องมือที่ใช้ในการทดสอบเว็บไซต์ โดยการจำลอง Browser ขึ้นมาโดยที่จะสามารถควบคุมเคอร์เซอร์เมาส์โดยอัตโนมัติ ใช้งานกับเว็บไซต์ที่มีการกดเข้าไปในตัวเว็บไซต์เพิ่มเติม
2. BeautifulSoup เป็น Python Module ที่ใช้สำหรับการดึงข้อมูล HTML กับ XML ออกมา เป็น library ในการทำ Web Scraping ที่ใช้งานง่ายและไม่ซับซ้อน

ทำ Web Scraping จากทั้งหมด 5 แหล่งข่าว และเลือกใช้แหล่งข่าวเพียง 3 แหล่งในการทำ WordCloud โดยมีรายละเอียดดังนี้

ตารางที่ 3-1 แสดงหัวข้อตารางโดยมีหัวข้อข่าวแหล่งที่มา เทคนิคที่ใช้ และการใช้ในการทำ WordCloud

ลำดับ	หัวข้อข่าวและแหล่งที่มา	เทคนิคที่ใช้	ใช้ในการทำ WordCloud	Code GitHub
1	ข้อมูลข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์ The Standard	Beautiful Soup	/	https://github.com/LUNA-eugene/DSI310_group2/

				blob/main/Analyze_News_1_Process/01_scraping.ipynb
2	ข้อมูลข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์ Thairath	Beautiful Soup Selenium	/	https://github.com/LUNA-eugene/DSI310_group2/blob/main/Analyze_News_2_Process/scraping_2.ipynb
3	ข้อมูลข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์ EEF (กองทุนเพื่อความเสมอภาค)	Beautiful Soup	/	https://github.com/LUNA-eugene/DSI310_group2/blob/main/WebScraping_Shopee_week3/scraping_3%20(1).ipynb
4	สำรวจราคาชุดนักเรียนของแต่ละจังหวัดใน Shopee	Beautiful Soup Selenium		https://github.com/LUNA-eugene/DSI310_group2/blob/main/News_Thairath_week4/scraping%20thairath%20week4.ipynb
5	ข่าวเกี่ยวกับสถานการณ์การศึกษาไทย จากเว็บไซต์ sanook	Beautiful Soup Selenium		https://github.com/LUNA-eugene/DSI310_group2/blob/main/WebScraping_5_Sanook.com/scraping%20sanook%20week%205.ipynb

อธิบายขั้นตอนการ Scraping

1. ทำการเลือกเว็บไซต์ที่จะทำการ Scrap ข่าวที่เกี่ยวข้องกับปัญหาของเด็กที่หลุดออกจากระบบการศึกษา
2. ทำการ Scrap ข่าวที่เราต้องการโดยใช้ library Selenium กับ BeautifulSoup
3. เมื่อทำการ Scrap เสร็จแล้ว ขั้นตอนต่อไปจะเป็นการนำข้อมูลมาสร้างเป็น Data Frame เพื่อนำมา Cleaning Data เช่น การเปลี่ยน Data type ของเวลาให้เป็น datetime โดยมี Format ให้ตรงตามมาตรฐาน ISO, การตัดคำที่ไม่เกี่ยวข้องออก เช่น Emoticon เป็นต้น
4. เมื่อ Cleaning Data แล้ว Export file เป็น CSV เพื่อนำข้อมูลไปใช้ทำ WordCloud

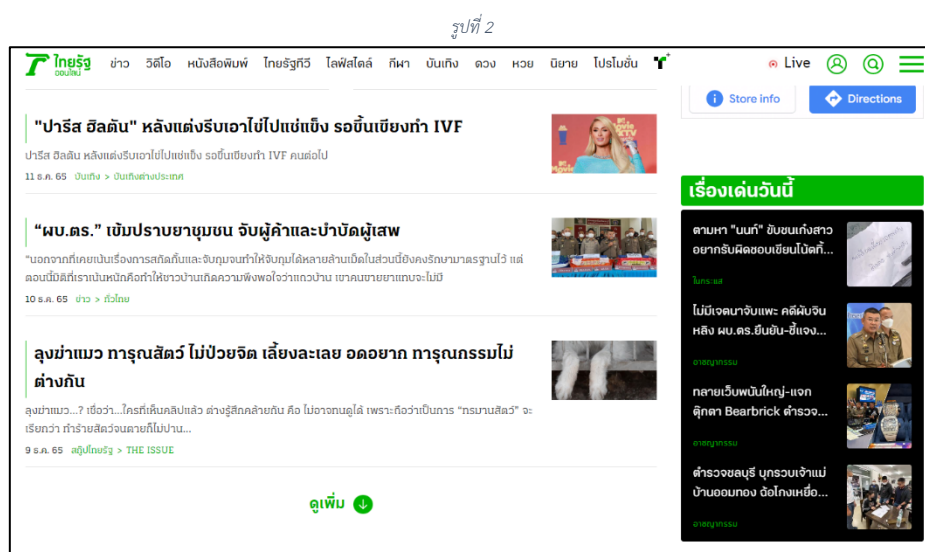
Code Scraping: GitHub - LUNA-eugene/DSI310_group2

ตัวอย่างการทำ Web Scraping

สำหรับตัวอย่างการทำ Web Scraping จากเว็บไซต์ Thairath เนื่องจากการใช้เครื่องมือในการ Scraping ทั้ง 2 เครื่องมือได้แก่ Selenium และ BeautifulSoup โดยมีรายละเอียดขั้นตอน ดังนี้

1. สืบค้นลักษณะของเว็บไซต์

เว็บไซต์ของไทยรัฐมีความพิเศษที่ต้องกดปุ่ม “ดูเพิ่ม” ถึงจะเห็นข่าวเพิ่มขึ้น ทำให้เราเลือกใช้ BeautifulSoup ในการ Scrap อย่างเดียวไม่ได้ จำเป็นต้องใช้ Selenium ในการควบคุมหน้าเว็บให้ป้อนคำสั่งพิมพ์ค่าในช่องค้นหา กดปุ่มเสร็จคำ และกดปุ่มดูเพิ่มเติม เราจึงเลือกใช้เครื่องมือ BeautifulSoup และ Selenium



รูปที่ 3-2 ภาพแสดงตัวอย่างหน้าเว็บไซต์ Thairath.com

2. Install และ import library ที่จำเป็นต้องใช้

รูปที่ 3

```
!pip install selenium

import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import bs4
import time
```

รูปที่ 3-3 ภาพแสดงการ import library

3. เครื่องมือ Selenium

3.1 การใช้เครื่องมือ Selenium ในการเปิดหน้าเว็บ Thairath ผ่านเว็บเบราว์เซอร์

3.2 ใช้คำสั่ง .find_element() ค้นหาปุ่มค้นหาผ่าน xpath และใช้คำสั่ง .click() ในการคลิกปุ่มค้นหา เพื่อจะป้อน input ในขั้นตอนถัดไป

3.3 ใช้คำสั่ง .find_element() ตาม xpath ที่ระบุเพื่อค้นหากล่องใส่ข้อมูล จากนั้นพิมพ์คำว่า “นักเรียนหลุดจากระบบ” ผ่านคำสั่ง send_keys()

รูปที่ 4

```
driver = webdriver.Chrome()
driver.get('https://www.thairath.co.th/home')

# คลิกเข้าสู่เว็บไซต์ไทยรัฐ
search = driver.find_element("xpath", '/html/body/div[2]/div/div/div[2]/a[2]')
search.click()

# คลิกปุ่มค้นหา
search = driver.find_element("xpath", '/html/body/div[2]/main/div/div[2]/div/div[1]/div[3]/div[2]/div[2]/button')
search.click()

# search คำว่า นักเรียนหลุดจากระบบ
searcher = driver.find_element("xpath", '/html/body/div[2]/main/div/div[2]/div/div[1]/div[3]/div[2]/div[2]/div/div/input')
searcher.send_keys('นักเรียนหลุดจากระบบ')
```

รูปที่ 3-4 ภาพแสดงตัวอย่างการใช้เครื่องมือ Selenium (1)

3.4 ยืนยันคำที่ต้องการจะค้นหาผ่านคำสั่ง .send_keys(keys_ENTER) และจะได้ผลลัพธ์เป็นลิสต์ข่าวของคำที่ต้องการค้นหา

3.5 ปรับขนาดหน้าให้เป็น 100% ผ่านคำสั่ง driver.execute_script("document.body.style.zoom='100%'")

3.6 จากนั้นจะทำการกดปุ่มดูเพิ่มเติมทั้งหมด 50 ครั้ง โดยการเขียน for loop และใช้คำสั่ง .find_element() ตาม xpath ของปุ่มดูเพิ่มเติม และใช้คำสั่ง .click() ในการคลิก

รูปที่ 5

```
from selenium.webdriver.common.keys import Keys

# ค้นหา
searcher.send_keys(Keys.ENTER)

driver.execute_script("document.body.style.zoom='100%'")

# คลิกปุ่มดูเพิ่มเติม 50 ครั้ง
for i in range(50):
    time.sleep(1)
    next_button = driver.find_element("xpath", '/html/body/div[2]/main/div/div[4]/div/div/div[1]/div[3]')
    next_button.click()
    time.sleep(1)
```

รูปที่ 3-5 ภาพแสดงตัวอย่างการใช้เครื่องมือ Selenium (2)

4. เครื่องมือ BeautifulSoup

4.1 เมื่อได้หน้าเว็บที่พร้อมสำหรับการ Scraping ต่อมาจะเป็นการใช้เครื่องมือ BeautifulSoup โดยจะทำการสร้าง list เพื่อเก็บข้อมูลก่อน

4.2 ใช้คำสั่ง driver.page_source เพื่อเก็บข้อมูลของ html ของเว็บไซต์ไว้ในตัวแปร data และสร้างตัวแปร soup เพื่อเก็บ object ของ BeautifulSoup

4.3 เขียน for loop และใช้คำสั่ง .find() ในการหา html หรือ css ที่ต้องการ เช่น ชื่อข่าว เนื้อข่าวโดยย่อ และวันที่เผยแพร่ข่าว แล้วใช้คำสั่ง .text() ในการสกัดเอาข้อความออกมาเก็บไว้ในตัวแปร title, content, date ตามลำดับ

4.4 นำตัวแปรผลลัพธ์ที่ได้เพิ่มเข้า list records ที่ประกาศไว้ตอนต้น โดยข่าวที่ได้จากเว็บไซต์ Thairath จะมีจำนวนทั้งสิ้น 255 ข่าว

รูปที่ 6

```

records = []

data = driver.page_source
soup = bs4.BeautifulSoup(data)

el = soup.find_all('div', class_='css-3douix e2foe9c4')

for e in el:
    title = e.find('h2', class_='css-11fy7ft efr6tej2').text.strip()
    content = e.find('p').text
    date = e.find('div', class_='css-1scjsja e2foe9c5').find('span').text

    records.append([title,content,date])

# จำนวนข่าว
x = len(records)
print(f'จำนวนข่าว: {x} ข่าว')

จำนวนข่าว: 255 ข่าว

```

รูปที่ 3-6 ภาพแสดงตัวอย่างการใช้เครื่องมือ BeautifulSoup

Prepare Data ของข้อมูลที่ได้จาก Web Scraping

1. นำข้อมูลที่ได้มาจัดเตรียมผ่านการทำ DataFrame

รูปที่ 7

```

# Create Data Frame
# Set Columns : title , content , date
df = pd.DataFrame(records, columns=['title','content','date'])
df

```

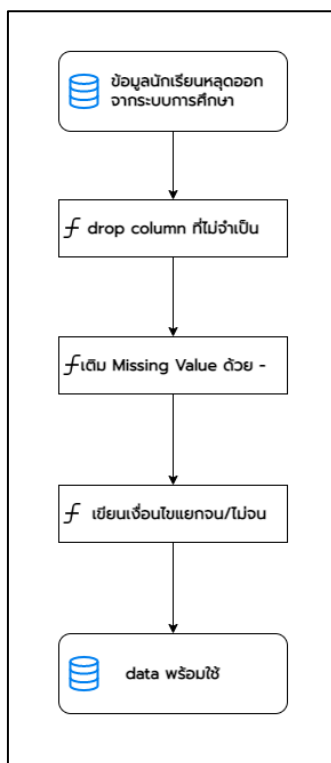
	title	content	date
0	พบเด็ก กศน. ค้างหวอดระบบขับแสบคน "รอลพ" ลี้...	น.ส.ศรีสุข เทียนทอง รว. ศักขาริการ ได้มีนโยบาย...	8 พ.ย. 65
1	สิ้นปีศึกษา 66 นร.ออกกลางคันเป็นศูนย์ "ตรีนุช"...	เปิดภาคเรียนที่ 2/2565 นี้ ศธ.มุ่งเน้นให้คุรุ...	2 พ.ย. 65
2	ความยากจนข้ามรุ่น การศึกษาเหลื่อมล้ำ	"ความยากจนข้ามรุ่น" และ "ความเหลื่อมล้ำทางการศ...	1 พ.ย. 65
3	ร่วม MOU เดินหน้า "ระบบหลักประกันความเสมอภาคทา...	ร่วม MOU เดินหน้า "ระบบหลักประกันความเสมอภาคทา...	31 ต.ค. 65
4	แนะหยุดวงจรความยากจนข้ามรุ่น ด้วยระบบโอกาสการศ...	แนะหยุดวงจรความยากจนข้ามรุ่น ด้วยการพัฒนาระบบ...	21 ต.ค. 65
...
250	พรณิษ ศก.เข้าเดิม กสศ.รุดช่วยเด็กยากจน เสี่ยง...	พรณิษ ศก.เข้า กสศ.รุดช่วยเด็กยากจน เสี่ยงหล...	5 มิ.ค. 63
251	ปูยาตัดโควิด-19 จากสอวกโกด หวาระบาดในไทย	ไทยพบผู้ติดเชื้อโควิด-19 เพิ่มขึ้นอีก 3 ร...	26 ก.พ. 63
252	"พรณิ" คุย "ลี จินกิง" สุวีรส มั่นใจจีนชน...	ประธานาธิบดีลี จินกิง ของจีน ยกพรณิศัพคย...	7 ก.พ. 63
253	ร.ร.ส่ง ค่อยอดเงินอุดหนุนรายหัว บริจาคให้ กสศ....	สวธิตเกษตร อินเคอร์ บุกเบิกค่อยอดเงินอุดหนุนร...	3 ก.พ. 63
254	กสศ.จับมือ กยศ.บูรณาการ ลดปัญหาให้ทุนศึกษาซ้ำ...	กสศ.จับมือ กยศ.บูรณาการฐานข้อมูล ลดปัญหาให้ทุน...	29 ม.ค. 63

รูปที่ 3-7 ภาพแสดงตัวอย่างข้อมูลที่ได้จาก Web Scraping

2. ทำการเปลี่ยน format ของ date จากเดิมที่เดือนเป็นตัวหนังสือและปีเป็น พ.ศ. โดยการเปลี่ยนเดือนเป็นตัวเลขและปีเป็นค.ศ. ผ่านคำสั่ง .replace() และใช้คำสั่ง .to_datetime() ในการเปลี่ยน format ปีเดือนวันที่ให้ถูกต้องตามมาตรฐาน และเปลี่ยน Data Type ให้เป็น datetime64

รูปที่ 3-11 ภาพแสดง Data pipeline ของข้อมูลสำหรับทำ Model & Dashboard

รูปที่ 12



รูปที่ 3-12 ภาพแสดง Data pipeline ของข้อมูลสำหรับการจำแนกเด็กยากจน/ไม่ยากจน

ตารางที่ 3-2 แสดงข้อมูล ประเภทของไฟล์ วัตถุประสงค์ และแหล่งที่มาของข้อมูลทุติยภูมิ

ลำดับ	ข้อมูล	ประเภทของไฟล์	วัตถุประสงค์	แหล่งที่มา
1	ข้อมูลเด็กหลุดออกจากระบบการศึกษา ระดับบุคคล (ก.ค. 65)	.xlsx	1. วิเคราะห์สาเหตุของเด็กหลุดออกจากระบบ 2. จำแนกเด็กที่มีฐานะยากจนกับไม่ยากจน	1. iSEE ฐานข้อมูลเพื่อความเสมอภาคทางการศึกษา 2. โครงการ “พาน้องกลับมาเรียน”
2	ข้อมูลจำนวนนักเรียนยากจน และยากจนพิเศษ จำแนกตามโรงเรียน ปีการศึกษา 2/2560 - 1/2564	.xlsx	1. เพื่อดูจำนวนและพื้นที่ของเด็กที่มีฐานะยากจนในการประกอบการพิจารณาเพื่อให้ทุนการศึกษา	iSEE ฐานข้อมูลเพื่อความเสมอภาคทางการศึกษา

3	รายได้เฉลี่ยต่อเดือนของ ครัวเรือน รายภาค และ รายจังหวัด ปี 2547 - 2564	.xlsx	1. นำมาสร้าง Model ทำนายจำนวนเด็กยากจน พิเศษและจำนวนเด็กที่ หลุดออกจากระบบ การศึกษา	สำนักงานสถิติ แห่งชาติ
4	หนี้สินเฉลี่ยต่อครัวเรือน จำแนกตามวัตถุประสงค์ ของการกู้ยืม เป็นราย จังหวัด พ.ศ. 2547 - 2564	.xlsx	1. นำมาสร้าง Model ทำนายจำนวนเด็กยากจน พิเศษและจำนวนเด็กที่ หลุดออกจากระบบ การศึกษา	สำนักงานสถิติ แห่งชาติ
5	งบประมาณด้าน การศึกษา ปีงบประมาณ 2565	.xlsx	1. ใช้ในการทำ Dashboard และโมเดล เพื่อวิเคราะห์สถานการณ์ ของเด็กที่หลุดออกจาก ระบบการศึกษา	ภาษีไปไหน? (ระบบข้อมูลการใช้ จ่ายภาครัฐ)

ตารางที่ 3

3.2. การเตรียมข้อมูลเบื้องต้น (Data Preprocessing / Data Cleaning)

3.2.1 ชุดข้อมูลที่ใช้สำหรับการทำ WordCloud

ใช้ข้อมูลทั้งหมดจากขั้นตอน Web Scraping ในการทำ WordCloud โดยการนำข้อมูลที่ได้จากแหล่งข้อมูลทั้ง 3 แหล่งข่าว นำมารวมให้อยู่ในไฟล์เดียวโดยใช้สกุลไฟล์ .xls ในการทำ Word Cloud library ที่เราเลือกใช้มีอยู่ 2 library ดังนี้

1. **PyThaiNLP** เป็น library ที่ใช้ในการวิเคราะห์และประมวลผลข้อมูลที่เป็นภาษาไทย โดยเฉพาะ
2. **WordCloud** เป็น library ที่ใช้ในการจับกลุ่มคำของข้อมูลโดยจะเรียงจากคำที่มีมากที่สุดไปคำที่มีน้อยที่สุด โดยจะแสดงผลให้ออกมาในรูปแบบ Visualization ให้มองเห็นคำที่มีมากที่สุด ได้ง่ายมากยิ่งขึ้น

ขั้นตอนการทำ Word Cloud

1. นำข้อมูลที่ได้จากการ Web Scraping ทั้ง 3 แหล่งข้อมูลนำมา รวมให้อยู่ในไฟล์เดียว
2. ใช้ฟังก์ชันจาก library ของ PyThaiNLP เพื่อในการสกัดคำโดยการตัดคำที่ไม่จำเป็นในการทำ WordCloud ออก

3. นำข้อมูลที่ผ่านการสกัดคำออกมาในรูปแบบของ Visualization ที่แสดงผลลัพธ์ของข้อมูลเป็นกลุ่มคำ หากเป็นคำที่ปรากฏเป็นจำนวนมากจะแสดงเป็นภาพคำขนาดใหญ่ แต่หากเป็นคำที่ปรากฏน้อยจะแสดงเป็นภาพคำขนาดเล็ก โดยการใช้ library WordCloud

3.2.2 ชุดข้อมูลจำนวนเด็กหลุดออกจากระบบการศึกษา

สำหรับชุดข้อมูลเด็กหลุดออกจากระบบการศึกษา (ก.ค. 2565) เบื้องต้นเราจะทำการลบนักเรียนที่ออกจากระบบด้วยสาเหตุการเสียชีวิตก่อน และลบคอลัมน์ที่ไม่ได้ใช้ เช่น ละติจูด ลองจิจูด รหัสโรงเรียน เป็นต้น

รูปที่ 13

สาเหตุหลัก	สาเหตุการออก	ความต้องการ	ความเห็นเพิ่มเติม
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	ความจำเป็นทางครอบครัว	ค่าใช้จ่ายในการสม	ผู้ปกครองพานักเรียนไป
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	นักเรียนจะเข้าเรียน กา
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	ผู้ปกครองมีรายได้น้อย ไม่พอเพียง	ให้คำแนะนำในการ	-
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	-
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	ความพิการ	บกพร่องทางการเรี	ด้านการเรียน ทางโรงเ
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	-
ไม่ประสงค์จะเรียนต่อ	อายุพ้นเกณฑ์และไม่ประสงค์จะเรียนต่อ	-	นักเรียนประสงค์ออกจา
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	-
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	-
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	ไปเรียนที่โรงเรียนเทศ
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	นร.อาศัยอยู่กับบิดา มา
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	นร. สมัครเรียนต่อที่ รร
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	-
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	ย้ายไปโรงเรียนเทศบา
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	การคมนาคมไม่สะดวก	-	ย้ายไปโรงเรียนบอสโก

รูปที่ 3-13 ภาพแสดงตัวอย่างข้อมูลสาเหตุการออกจากระบบการศึกษาของนักเรียน

จากนั้น หากพิจารณาถึงสาเหตุการออกจากระบบของนักเรียนเพียงแค่ข้อมูลสาเหตุหลักและสาเหตุการออกอาจจะทำให้มองข้ามปัญหาเรื่องความจนที่แฝงมาในข้อมูลความต้องการและความคิดเห็นเพิ่มเติมได้ เช่น

1. ข้อมูลระบุสาเหตุการออก คือ ความจำเป็นของครอบครัว แต่ในด้านความต้องการระบุถึงความต้องการทุนการศึกษา ซึ่งสื่อถึงปัญหาด้านความจน
2. ข้อมูลระบุสาเหตุการออก คือ ความจำเป็นทางครอบครัว แต่ในความคิดเห็นเพิ่มเติมระบุถึงความต้องการออกไปทำงานช่วยเหลือครอบครัว ซึ่งสื่อถึงปัญหาด้านความจน เป็นต้น

รูปที่ 14

สาเหตุหลัก	สาเหตุการออก	
ยังไม่กลับเข้าระบบ (จำแนกตามสาเหตุ)	ผู้ปกครองมีรายได้น้อย ไม่พอเพียง	568
	ความจำเป็นทางครอบครัว	556
	ผู้รายงานไม่ได้ระบุ	350
	ย้ายถิ่นที่อยู่	310
	สภาพของครอบครัว	256
	ความพิการ	161
	ปัญหาการปรับสภาพตัวเองในสังคม	115
	สุขภาพอนามัย	81
	ได้รับผลกระทบจาก Covid-19	47
	การประพฤติ ปฏิบัติขัดกับระเบียบของสถานศึกษา	29
	การคมนาคมไม่สะดวก	22
	อยู่ในกลุ่มเสี่ยง	22
	ได้รับผลกระทบจากภัยต่างๆ	4
ไม่ประสงค์จะเรียนต่อ	จบการศึกษามาถึงระดับและไม่ประสงค์จะเข้าเรียนต่อ	2234
	อายุพ้นเกณฑ์และไม่ประสงค์จะเรียนต่อ	543

รูปที่ 3-14 ภาพแสดงข้อมูลสรุปสาเหตุหลักและสาเหตุการออก

ทำให้ต้องมีการเตรียมข้อมูลโดยสร้างเป็นคอลัมน์ใหม่ที่สามารถคาดการณ์ได้ว่านักเรียนคนนี้มีภาวะความยากจนหรือไม่ เพื่อสื่อถึงสาเหตุที่แท้จริงที่ทำให้เด็กหลุดออกจากระบบการศึกษา โดยจะใช้การเขียนเงื่อนไขที่จะพิจารณาภาวะความยากจนของนักเรียนโดยดูจาก Keywords ต่าง ๆ โดยมีลำดับของเงื่อนไข ดังนี้

รูปที่ 15

```
# Create Condition to find จอ
conditions = [(df['สาเหตุการออก'] == 'ผู้ปกครองมีรายได้น้อย ไม่พอเพียง') |
(df['ความต้องการ'].str.contains('ทุน|เงิน|จน|รายได้อาชีพ|ทำงาน|ประกอบอาชีพคนเดียว|ค่าใช้จ่าย|งาน|งบประมาณ|ลำบาก|ช่วยเหลือในด้านปัจจัยพื้นฐาน|ต่อ
(df['ความเห็นเพิ่มเติม'].str.contains('ทุน|เงิน|จน|รายได้อาชีพ|ทำงาน|ประกอบอาชีพคนเดียว|ค่าใช้จ่าย|ไม่มั่นคง|ขาดแคลน|ช่วยเหลือ|ลำบาก|รับจ้าง|ช่วยครอบครัว
```

รูปที่ 3-15 ภาพแสดงตัวอย่างโค้ดการเขียนเงื่อนไข

เงื่อนไขที่ 1 สาเหตุการออก

หากสาเหตุการออก คือ "ผู้ปกครองมีรายได้น้อย ไม่พอเพียง" แสดงว่าเด็กคนนี้มีภาวะความยากจน แต่นอกจากสาเหตุการออกนี้ ในสาเหตุอื่น ๆ ก็สามารถมีภาวะความยากจนแฝงได้เช่นกัน ดังนั้นจึงพิจารณาในเงื่อนไขถัดไป

เงื่อนไขที่ 2 ความต้องการ

หากนักเรียนระบุ Keywords ความต้องการที่สื่อถึงความยากจน เช่น ต้องการทุนการศึกษา, ความลำบาก หรือ ความยากจน แสดงว่าเด็กคนนี้มีภาวะความยากจน ซึ่งหากไม่มี Keywords เหล่านี้จะให้พิจารณาในเงื่อนไขที่ 3 ต่อไป

เงื่อนไขที่ 3 ความเห็นเพิ่มเติม

มี Keywords ที่สื่อถึงความยากจน เช่น ขาดแคลน, ไม่มั่นคง, ทำไร่นานา, เลี้ยงวัว เป็นต้น แสดงว่าเด็กคนนี้มีภาวะยากจน และหากไม่เข้าเงื่อนไขทั้งหมดก็แสดงว่าข้อมูลอาจไม่เพียงพอจึงคาดว่าไม่ยากจน

3.2.3 ชุดข้อมูลสำหรับการทำ Model

ข้อมูลที่นำมาใช้สำหรับการทำ Model ได้แก่ ภาค, จำนวนเด็กหลุดออกจากระบบ, รายได้เฉลี่ยต่อครัวเรือน, หนี้สินเฉลี่ยต่อครัวเรือน, หนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน, งบประมาณด้านการศึกษา และจำนวนเด็กยากจนพิเศษของแต่ละจังหวัด

ขั้นตอนการทำชุดข้อมูลสำหรับการทำ Model

1. ทำการรวมจำนวนเด็กหลุดออกจากระบบการศึกษาของแต่ละจังหวัด ในชุดข้อมูลจำนวนเด็กหลุดออกจากระบบการศึกษา
2. หาหนี้สินเฉลี่ยต่อครัวเรือน โดยการนำข้อมูลหนี้สินทั้งหมดต่อครัวเรือน มาลบกับข้อมูลหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน
3. นำข้อมูลที่ต้องใช้ในการทำ Model มารวมกันให้อยู่ในไฟล์เดียว
4. เพิ่มคอลัมน์ภาคเข้าไป เพื่อใช้ในการวิเคราะห์แทนคอมมูนิตี้จังหวัด

3.3. การวิเคราะห์ภาพรวมข้อมูล (Exploratory Data Analysis / Data Visualization)

3.3.1 WordCloud

สำหรับข้อมูลที่นำมาใช้ในการทำ WordCloud เป็นข้อมูลของแหล่งข่าวทั้งหมดที่รวมอยู่ในไฟล์เดียวกันโดยจะวิเคราะห์ความเชื่อมโยงหรือ Keywords ที่มีความสำคัญและเกี่ยวข้องกับปัญหาเพื่อหาว่าปัจจัยใดที่ส่งผลทำให้เด็กหลุดออกจากระบบการศึกษาได้มากที่สุด

3.3.2 Dashboard และ Model

สำหรับข้อมูลที่ใช้ในการทำ Dashboard จะประกอบไปด้วยชุดข้อมูลจำนวน 4 ชุด ได้แก่ ข้อมูลเด็กหลุดออกจากระบบการศึกษา ระดับบุคคล (ก.ค. 65), ข้อมูลจำนวนนักเรียนยากจนและยากจนพิเศษ จำแนกตามโรงเรียน, รายได้เฉลี่ยต่อเดือนของครัวเรือน รายภาค และรายจังหวัด ปี 2564 และหนี้สินเฉลี่ยต่อครัวเรือน จำแนกตามวัตถุประสงค์ของการกู้ยืม เป็นรายจังหวัด พ.ศ. 2547 - 2564 โดยมีจุดประสงค์เพื่อดูว่าพื้นที่ไหนต้องการความช่วยเหลือในด้านทุนทรัพย์มากที่สุดและเป็นแนวทางในการมอบทุนให้เด็กในแต่ละพื้นที่ต่อไป

รูปที่ 16

	บร.ยากจนพิเศษ	จำนวนเด็กหลุดออกจากระบบ (คน)	รายได้เฉลี่ยต่อครัวเรือน (บาท)	หนี้สินเฉลี่ยต่อครัวเรือน (บาท)	หนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน (บาท)	งบประมาณด้านการศึกษา (ล้านบาท)
count	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000
mean	15693.077922	68.805195	24666.334286	199463.175714	3483.761429	5894.701299
std	17267.348759	100.992970	5762.322093	73320.706375	2679.435627	33697.422150
min	412.000000	0.000000	15495.950000	47602.030000	0.000000	236.000000
25%	3283.000000	3.000000	20691.690000	151192.640000	1547.000000	593.000000
50%	8132.000000	30.000000	23866.820000	194002.850000	2986.100000	1318.000000
75%	25367.000000	98.000000	26921.920000	238611.200000	4594.410000	2156.000000
max	72133.000000	469.000000	41129.420000	367899.620000	14265.750000	296911.000000

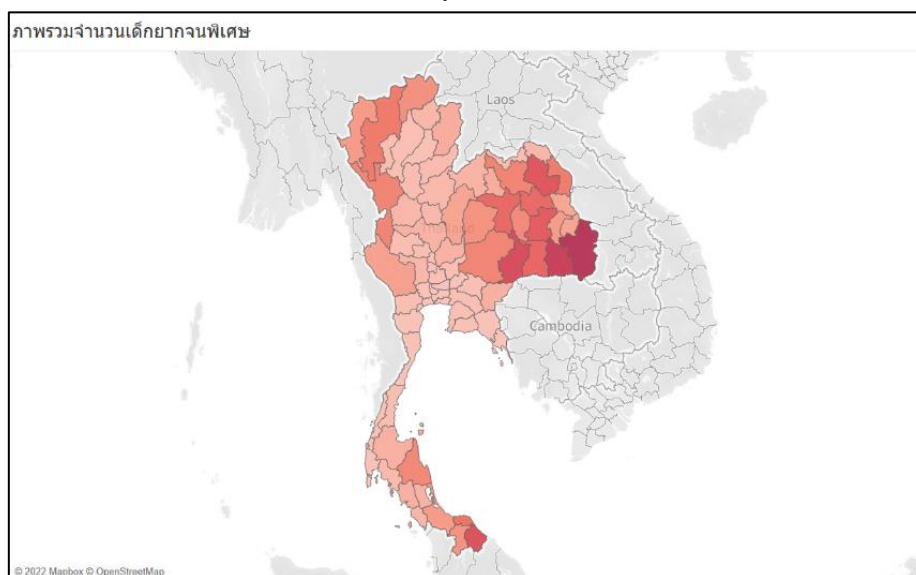
รูปที่ 3-16 ภาพแสดงภาพรวมของข้อมูลที่ใช้ในการทำโมเดล

จากรูปที่ 3-16 จะเห็นว่าค่าเฉลี่ยของจำนวนนักเรียนยากจนพิเศษของแต่ละจังหวัดจะอยู่ที่ 15,693 คน, ค่าเฉลี่ยของเด็กหลุดออกจากระบบของแต่ละจังหวัดอยู่ที่ 69 คน, ค่าเฉลี่ยของรายได้เฉลี่ยต่อครัวเรือนในแต่ละจังหวัด คือ 24,666 บาท, หนี้สินต่อครัวเรือนเมื่อหักหนี้สินเพื่อการศึกษาเฉลี่ยอยู่ที่ 199,463 บาท, หนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือนของแต่ละจังหวัด คือ 3,483 บาท และงบประมาณด้านการศึกษาแต่ละจังหวัดได้เฉลี่ย 5,894 ล้านบาท

3.2.3 ภาพรวมข้อมูลทั้งหมด

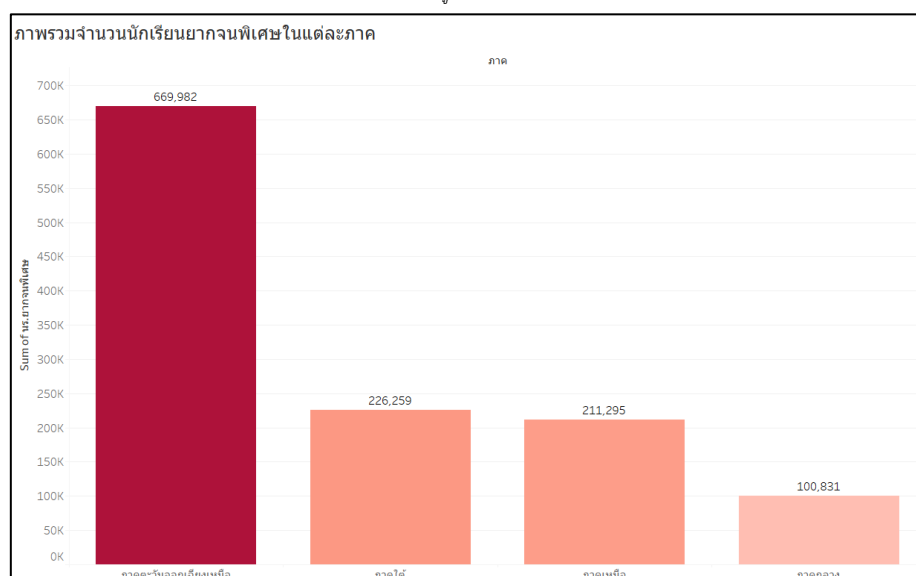
ข้อมูลจำนวนเด็กยากจนพิเศษ

รูปที่ 17



รูปที่ 3-17 ภาพแสดงภาพรวมจำนวนเด็กยากจนพิเศษ

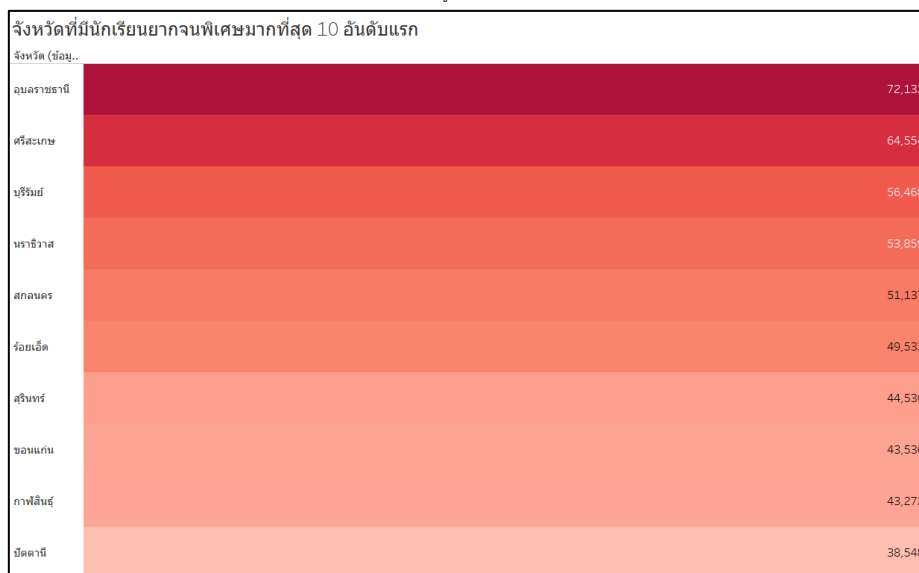
รูปที่ 18



รูปที่ 3-18 ภาพแสดงภาพรวมจำนวนเด็กยากจนพิเศษในแต่ละภาค

จากรูปที่ 3-17 และ 3-18 จะสังเกตได้ว่า จังหวัดที่มีนักเรียนยากจนพิเศษเยอะเป็นพิเศษ ส่วนใหญ่จะเป็นจังหวัดที่อยู่ตามชายแดนของประเทศ โดยภาคที่มีจำนวนนักเรียนยากจนพิเศษมากที่สุด คือ ภาคตะวันออกเฉียงเหนือ ภาคใต้ ภาคเหนือ และภาคกลาง ตามลำดับ

รูปที่ 19

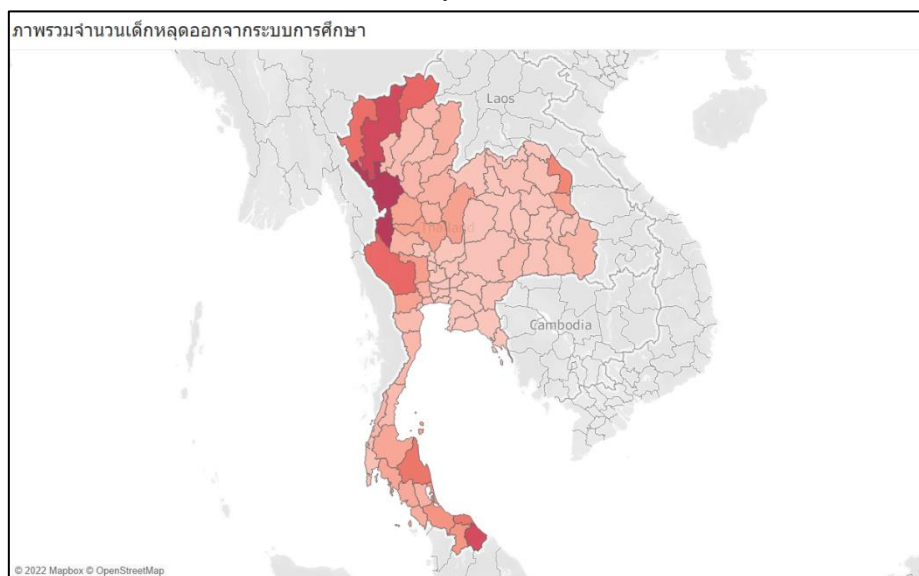


รูปที่ 3-19 ภาพแสดงจังหวัดที่มีนักเรียนยากจนพิเศษมากที่สุด 10 อันดับแรก

จากรูปที่ 3-19 จังหวัดที่มีจำนวนนักเรียนยากจนพิเศษสูงที่สุด คือ จังหวัดอุดรธานี มีจำนวนนักเรียนยากจนพิเศษมากถึง 72,133 คน เมื่อเปรียบเทียบกับจังหวัดที่มีจำนวนนักเรียนยากจนพิเศษน้อยที่สุดอย่างจังหวัดสมุทรสงคราม จำนวน 412 คน จะเห็นว่า จังหวัดอุดรธานีมีจำนวนนักเรียนยากจนพิเศษคิดเป็น 175 เท่าของจังหวัดสมุทรสงคราม และสังเกตได้ว่าจังหวัดที่มีนักเรียนยากจนพิเศษมากที่สุดส่วนใหญ่จะเป็นจังหวัดนภาคตะวันออกเฉียงเหนือ

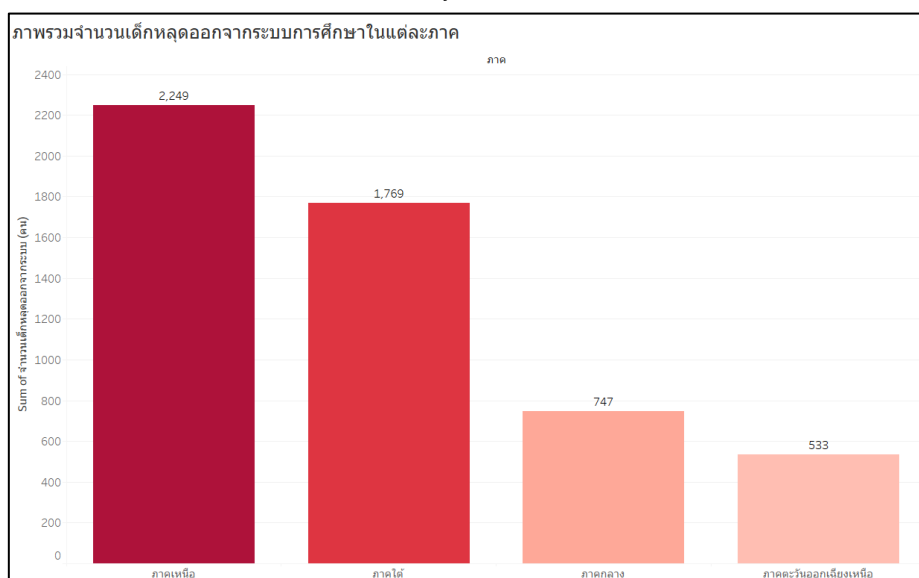
ข้อมูลจำนวนเด็กหลุดออกจากระบบการศึกษา

รูปที่ 20



รูปที่ 3-20 ภาพแสดงจำนวนเด็กที่หลุดออกจากระบบการศึกษา

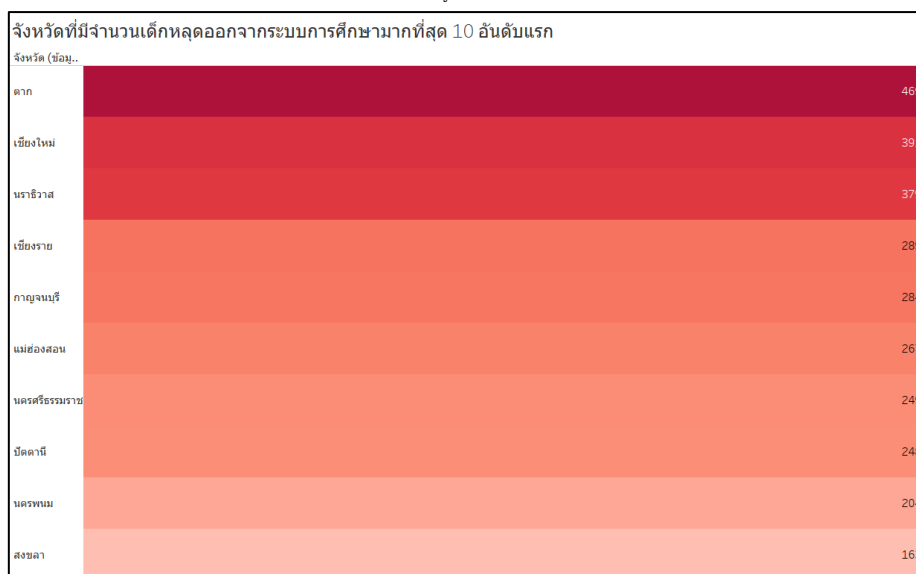
รูปที่ 21



รูปที่ 3-21 ภาพแสดงภาพรวมจำนวนเด็กหลุดออกจากระบบการศึกษาในแต่ละภาค

จากรูปที่ 3-20 และ 3-21 เด็กที่หลุดออกจากระบบการศึกษาในเดือนกรกฎาคม พ.ศ. 2565 ส่วนใหญ่จะเป็นเด็กที่อาศัยอยู่ในภาคเหนือ ภาคใต้ ภาคกลาง และภาคตะวันออกเฉียงเหนือ ตามลำดับ

รูปที่ 22

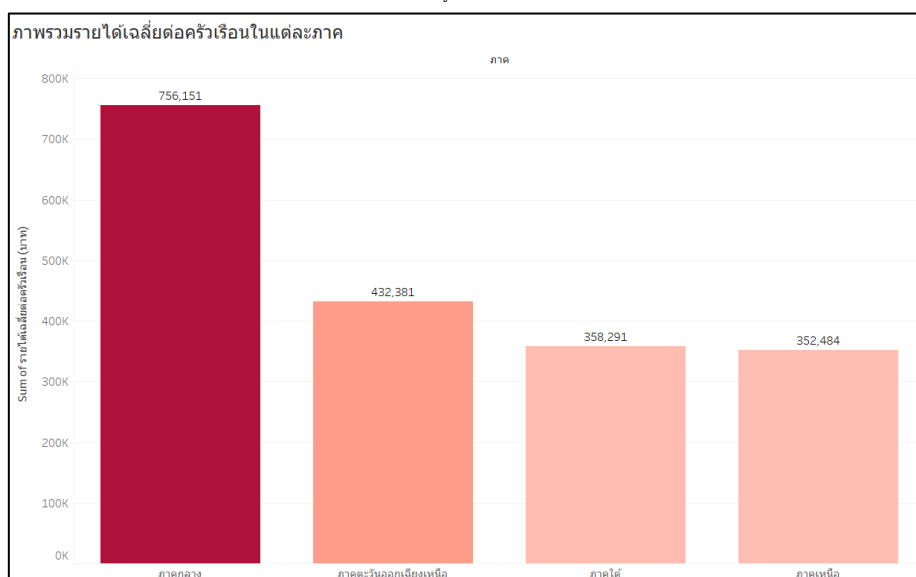


รูปที่ 3-22 ภาพแสดงจังหวัดที่มีเด็กหลุดออกจากระบบการศึกษา 10 อันดับ

จากรูปที่ 3-22 จังหวัดที่มีจำนวนนักเรียนหลุดออกจากระบบมากที่สุดในเดือนกรกฎาคม พ.ศ. 2565 คือ จังหวัดตาก โดยมีจำนวนนักเรียนหลุดออกจากระบบทั้งหมด 469 คน และจังหวัดที่มีจำนวนนักเรียนหลุดออกจากระบบการศึกษาส่วนใหญ่เป็นจังหวัดในภาคเหนือและภาคใต้เป็นหลัก

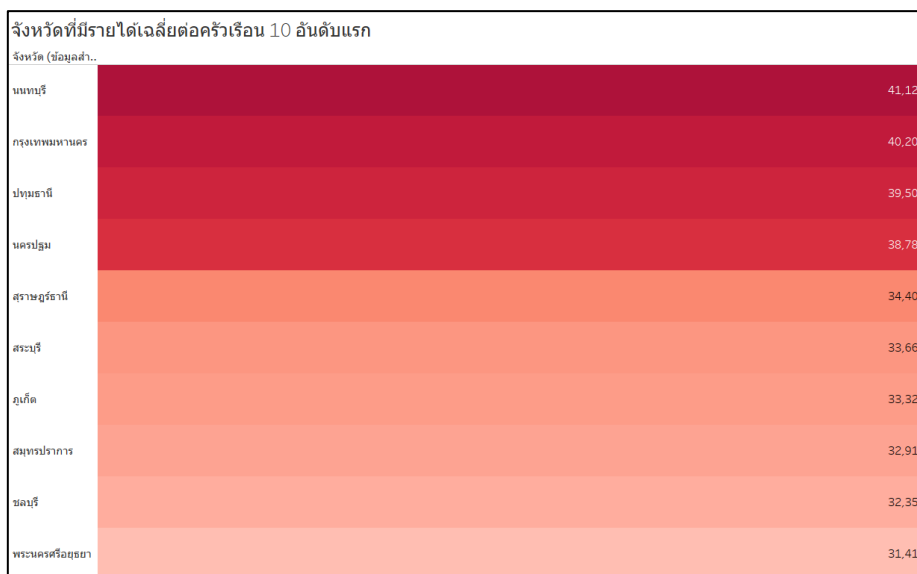
ข้อมูลรายได้เฉลี่ยต่อครัวเรือน

รูปที่ 23



รูปที่ 3-23 ภาพแสดงรายได้เฉลี่ยต่อครัวเรือนในแต่ละภาค

รูปที่ 24



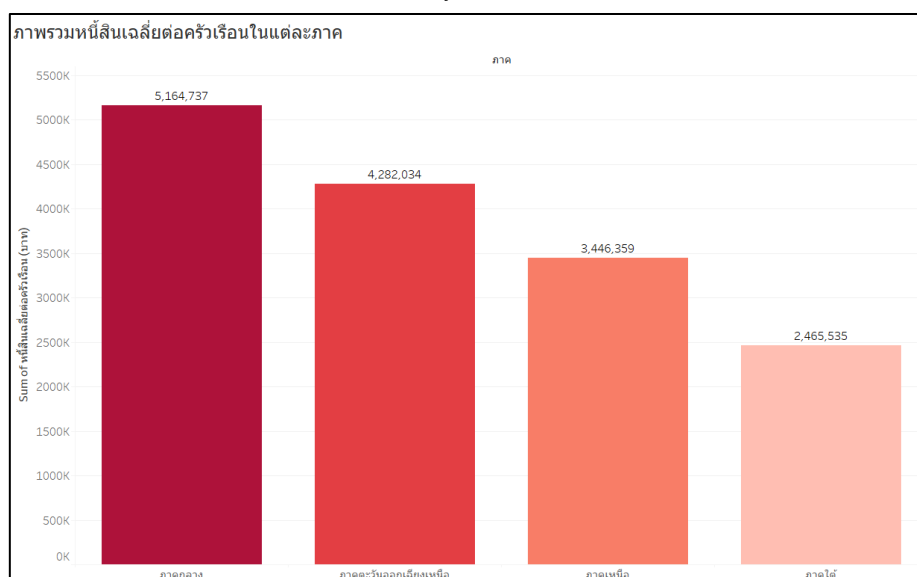
รูปที่ 3-24 ภาพแสดงรายได้เฉลี่ยต่อครัวเรือนในแต่ละจังหวัด

จากรูปที่ 3-23 และ 3-24 สำหรับรายได้เฉลี่ยต่อครัวเรือน จะเห็นว่า ภาคกลางมีรายได้เฉลี่ยต่อครัวเรือนสูงที่สุด รองลงมาภาคตะวันออกเฉียงเหนือ ภาคใต้ และภาคเหนือ ตามลำดับ นอกจากนี้ยังสังเกตได้อีกว่า ภาคที่มีรายได้เฉลี่ยต่อครัวเรือนมากที่สุดอย่าง “ภาคกลาง” มีรายได้เฉลี่ยมากกว่า “ภาคเหนือ” ซึ่งเป็นภาคที่มีรายได้เฉลี่ยต่อครัวเรือนน้อยที่สุดถึง 2 เท่า

และอีกหนึ่งสิ่งที่สามารถสังเกตได้ คือ ภาคตะวันออกเฉียงเหนือซึ่งมีรายได้เฉลี่ยต่อครัวเรือนมากที่สุดเป็นอันดับ 2 กลับมีจำนวนนักเรียนยากจนพิเศษเป็นอันดับ 1 ตรงส่วนนี้ทำให้สามารถสันนิษฐานได้ว่า “ภาคตะวันออกเฉียงเหนือเป็นภาคที่มีความเหลื่อมล้ำสูงที่สุด”

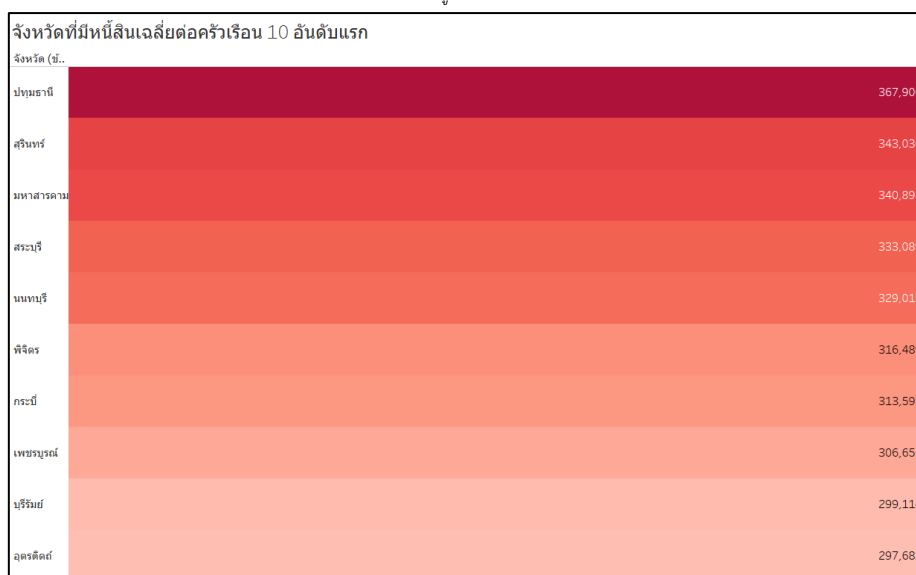
ข้อมูลหนี้สินเฉลี่ยต่อครัวเรือน

รูปที่ 25



รูปที่ 3-25 ภาพแสดงหนี้สินเฉลี่ยต่อครัวเรือนในแต่ละภาค

รูปที่ 26

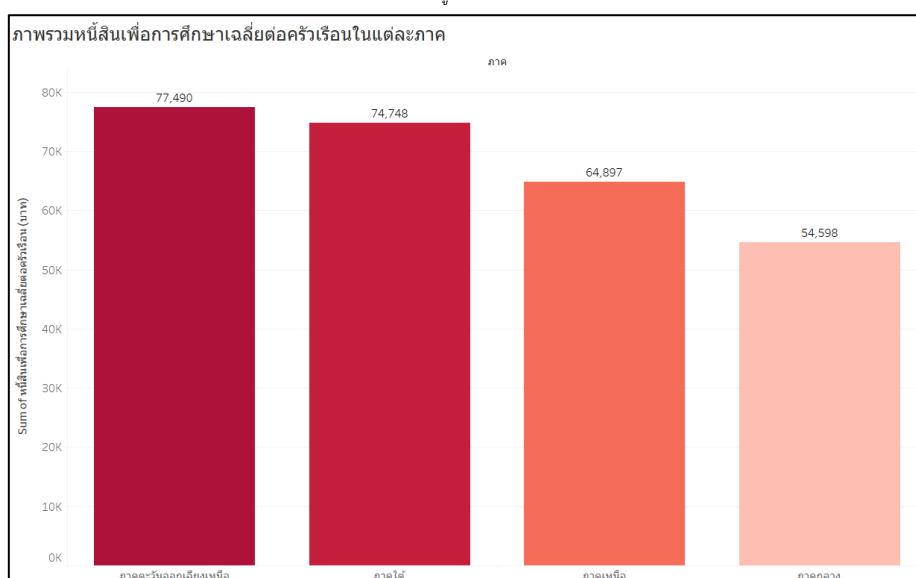


รูปที่ 3-26 ภาพแสดงหนี้สินเฉลี่ยต่อครัวเรือนในแต่ละจังหวัด

จากรูปที่ 3-25 และ 3-26 จะเห็นว่า ภาคที่มีจำนวนหนี้สินเฉลี่ยต่อครัวเรือนมากที่สุด คือ ภาคกลาง ภาคตะวันออกเฉียงเหนือ ภาคเหนือ และภาคใต้ โดยข้อมูลจะเป็นไปตามรายได้เฉลี่ยต่อครัวเรือนว่าหากมีหนี้สินเฉลี่ยต่อครัวเรือนสูง รายได้เฉลี่ยต่อครัวเรือนก็จะสูงตามไปด้วย และจังหวัดที่มีหนี้สินเฉลี่ยต่อครัวเรือนมากที่สุด คือ จังหวัดปทุมธานี ซึ่งมีหนี้สินเฉลี่ยต่อครัวเรือนทั้งสิ้น 367,900 บาท

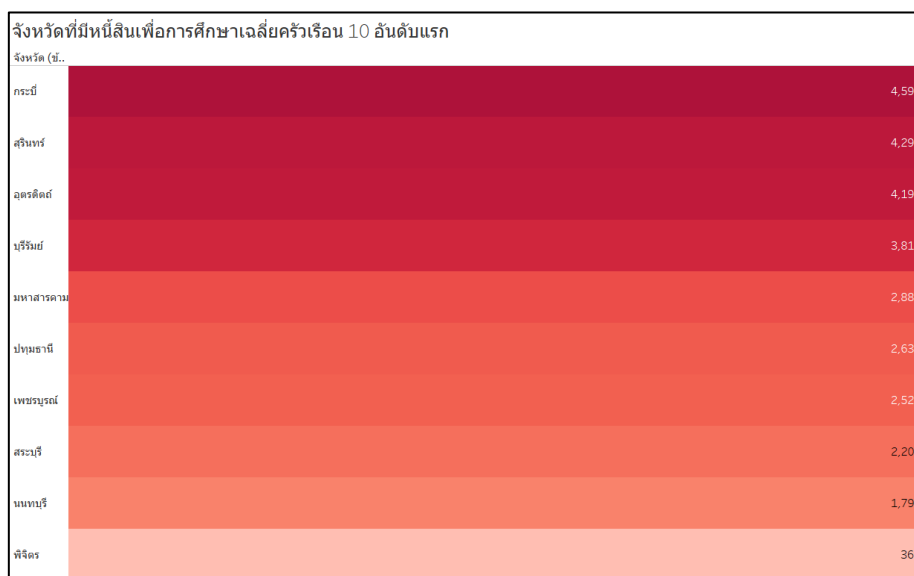
ข้อมูลหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน

รูปที่ 27



รูปที่ 3-27 ภาพแสดงหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือนในแต่ละภาค

รูปที่ 28

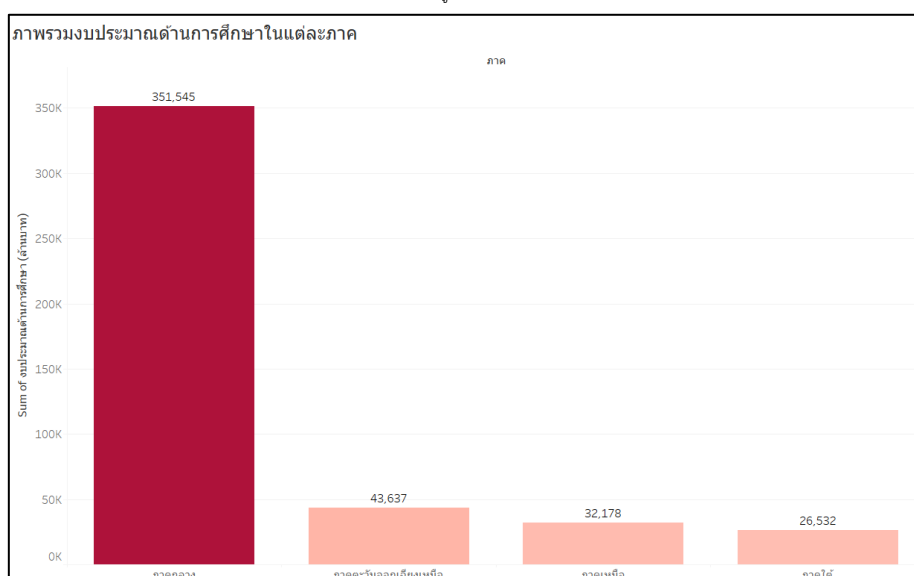


รูปที่ 3-28 ภาพแสดงหนี้สิ้นเพื่อการศึกษาเฉลี่ยต่อครัวเรือนในแต่ละจังหวัด

จากรูปที่ 3-27 และ 3-28 จะเห็นว่า ภาคที่มีหนี้สิ้นเพื่อการศึกษาเฉลี่ยต่อครัวเรือนมากที่สุด คือ ภาคตะวันออกเฉียงเหนือ ภาคใต้ ภาคเหนือ และภาคกลาง โดยข้อมูลจะเป็นไปในทิศทางเดียวกับ จำนวนนักเรียนยากจนพิเศษที่ถ้าหากมีจำนวนหนี้สิ้นเพื่อการศึกษาสูง จำนวนนักเรียนยากจนพิเศษก็จะสูงตามไปด้วย และจังหวัดที่มีหนี้สิ้นเพื่อการศึกษาเฉลี่ยต่อครัวเรือนมากที่สุด คือ จังหวัดกระบี่ ซึ่งมีหนี้สิ้นเพื่อการศึกษาเฉลี่ยต่อครัวเรือนทั้งสิ้น 4,594 บาท

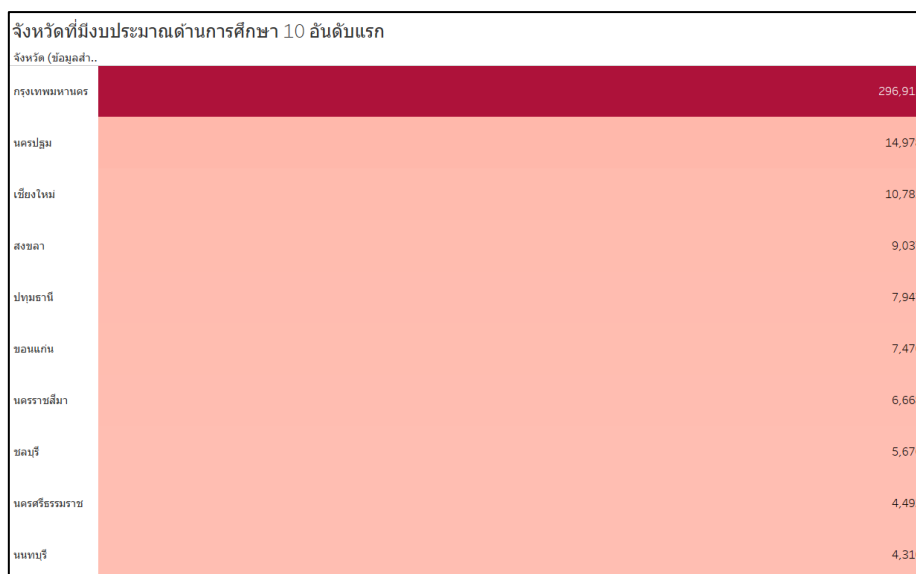
ข้อมูลงบประมาณด้านการศึกษา

รูปที่ 29



รูปที่ 3-29 ภาพแสดงงบประมาณด้านศึกษาในแต่ละภาค

รูปที่ 30



รูปที่ 3-30 ภาพแสดงงบประมาณด้านศึกษาในแต่ละจังหวัด

จากรูปที่ 3-29 และ 3-30 จะเห็นว่า ภาคที่มีงบประมาณการศึกษามากที่สุด คือ ภาคกลาง ภาคตะวันออกเฉยงเหนือ ภาคเหนือ และภาคใต้ ตามลำดับ ซึ่งจะสามารถสังเกตได้ว่า หากเป็นภาคที่ได้รับงบประมาณด้านการศึกษาต่ำจะมีจำนวนนักเรียนยากจนพิเศษสูง

จังหวัดที่มีงบประมาณด้านการศึกษามากที่สุด คือ จังหวัดกรุงเทพมหานคร โดยได้รับงบประมาณด้านการศึกษามากกว่าจังหวัดอื่นและจังหวัดที่ได้รับงบประมาณน้อยที่สุดอย่างจังหวัดสมุทรสงครามมากถึง 1,258 เท่า นอกจากนี้ยังสังเกตได้ว่า จังหวัดที่ได้รับงบประมาณด้านการศึกษามากที่สุด 10 อันดับส่วนใหญ่มักเป็นเมืองใหญ่

3.4. การสร้างแบบจำลอง (Model Building)

ทำการสร้าง Model จำนวน 2 Model ได้แก่

1. Model เพื่อทำนายจำนวนนักเรียนที่หลุดออกจากระบบการศึกษา
2. Model เพื่อทำนายจำนวนนักเรียนยากจนพิเศษ

โดยรายละเอียดของชุดข้อมูลที่ใช้ประกอบการทำ Model มีดังนี้

ตารางที่ 3-3 รายละเอียดของชุดข้อมูลที่จะใช้แต่ละ Model

ลำดับ	ชื่อคอลัมน์	คำอธิบาย	Data Type	Model 1	Model 2
1	ภูมิภาค	ภูมิภาคตามจังหวัด	object	feature (x1)	feature (x1)

2	นร.ยากจนพิเศษ	จำนวนนักเรียนยากจนพิเศษ* ปีการศึกษา 1/2564 (* นักเรียนยากจนพิเศษคือ ครอบครัวที่มีรายได้เฉลี่ย 1,200 บาทต่อคนต่อเดือน)	int64	feature (x2)	label (y)
3	จำนวนเด็กหลุดออกจากระบบ (คน)	จำนวนเด็กหลุดออกจากระบบการศึกษา ก.ค.2565	int32	label (y)	feature+ ++ (x2)
4	รายได้เฉลี่ยต่อครัวเรือน (บาท)	รายได้เฉลี่ยต่อครัวเรือน (บาท) ปี 2564	float64	feature (x3)	feature (x3)
5	หนี้สินเฉลี่ยต่อครัวเรือน (บาท)	หนี้สินเฉลี่ยต่อครัวเรือน* (บาท) ปี 2564 (*หนี้สินเฉลี่ยต่อครัวเรือน = หนี้สินเฉลี่ยต่อครัวเรือนด้วยสาเหตุทั้งหมด - หนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน)	float64	feature (x4)	feature (x4)
6	หนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน (บาท)	หนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือน (บาท) ปี 2564	float64	feature (x5)	feature (x5)
7	งบประมาณด้านการศึกษา (ล้านบาท)	งบประมาณด้านการศึกษา (ล้านบาท) ปีงบประมาณ 2565	int64	feature (x6)	feature (x6)

ตารางที่ 4

สำหรับการสร้าง Model จะเลือกใช้ **Pycaret** ซึ่งเป็น open-source, Low-code Machine Learning library ที่จะช่วยให้สามารถสร้าง Model ได้สะดวกรวดเร็วมากขึ้น ซึ่งอีกหนึ่งจุดเด่นของ Pycaret คือ ช่วยให้สามารถเปรียบเทียบผลลัพธ์จาก Machine Learning library ต่าง ๆ เช่น Linear Regression, AdaBoost, lightgbm เป็นต้น นอกจากนี้ ยังสามารถในการ Preprocessing Data และ Optimize Model ได้อีกด้วย

รูปที่ 31



รูปที่ 3-31 ภาพโลโก้ Pycaret และ Features ที่สำคัญ

รูปที่ 32

ID	Name	Reference	Turbo
lr	Linear Regression	sklearn.linear_model._base.LinearRegression	True
lasso	Lasso Regression	sklearn.linear_model._coordinate_descent.Lasso	True
ridge	Ridge Regression	sklearn.linear_model._ridge.Ridge	True
en	Elastic Net	sklearn.linear_model._coordinate_descent.Elast...	True
lar	Least Angle Regression	sklearn.linear_model._least_angle.Lars	True
llar	Lasso Least Angle Regression	sklearn.linear_model._least_angle.LassoLars	True
omp	Orthogonal Matching Pursuit	sklearn.linear_model._omp.OrthogonalMatchingPu...	True
br	Bayesian Ridge	sklearn.linear_model._bayes.BayesianRidge	True
ard	Automatic Relevance Determination	sklearn.linear_model._bayes.ARDRidge	False
par	Passive Aggressive Regressor	sklearn.linear_model._passive_aggressive.Passi...	True
ransac	Random Sample Consensus	sklearn.linear_model._ransac.RANSACRegressor	False
tr	TheilSen Regressor	sklearn.linear_model._theil_sen.TheilSenRegressor	False
huber	Huber Regressor	sklearn.linear_model._huber.HuberRegressor	True
kr	Kernel Ridge	sklearn.kernel_ridge.KernelRidge	False
svm	Support Vector Regression	sklearn.svm._classes.SVR	False
knn	K Neighbors Regressor	sklearn.neighbors._regression.KNeighborsRegressor	True
dt	Decision Tree Regressor	sklearn.tree._classes.DecisionTreeRegressor	True
rf	Random Forest Regressor	sklearn.ensemble._forest.RandomForestRegressor	True
et	Extra Trees Regressor	sklearn.ensemble._forest.ExtraTreesRegressor	True
ada	AdaBoost Regressor	sklearn.ensemble._weight_boosting.AdaBoostRegr...	True
gbr	Gradient Boosting Regressor	sklearn.ensemble._gb.GradientBoostingRegressor	True
mip	MLP Regressor	sklearn.neural_network._multilayer_perceptron....	False
lightgbm	Light Gradient Boosting Machine	lightgbm.sklearn.LGBMRegressor	True
dummy	Dummy Regressor	sklearn.dummy.DummyRegressor	True

รูปที่ 3-32 Machine Learning library สำหรับการทำให้ Regression ที่ Pycaret รองรับ

สำหรับขั้นตอนการสร้าง Model มีดังนี้

3.4.1 Normalization: Robust Scaler

Normalization หรือ Feature Scaling คือ การปรับช่วงข้อมูลของแต่ละ Feature ให้เป็นช่วงเดียวกันก่อนจะนำไปใช้งานจึงเป็นขั้นตอนที่มีความสำคัญมาก เพราะ ข้อมูลที่นำมาใช้มีความหลากหลายในด้าน Scale ซึ่งบาง Model ยังไม่รองรับตรงจุดนี้จึงต้องทำการปรับ Scale ของแต่ละ Feature ให้อยู่ในช่วงเดียวกันก่อน โดยเทคนิคที่นำมาใช้ คือ เทคนิค Robust Scaler คือ การ Scale เพื่อกำจัด Outlier โดยจะใช้ค่า Median กับ Quantile แทน

รูปที่ 33

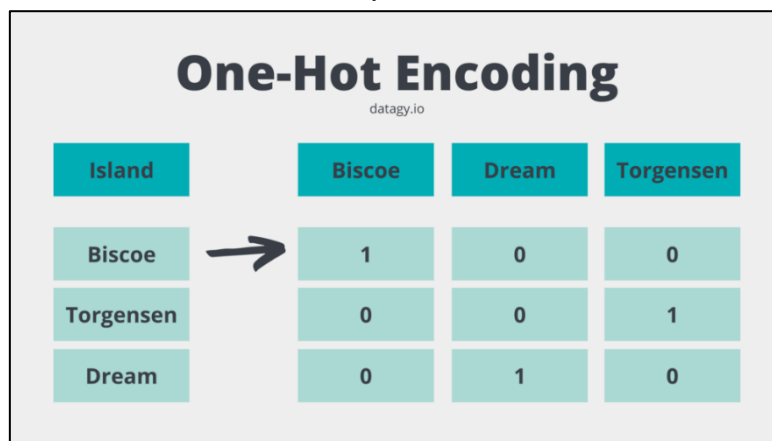
$$X_{\text{scale}} = \frac{x_i - x_{\text{med}}}{x_{75} - x_{25}}$$

รูปที่ 3-33 ภาพสูตรการคำนวณ Robust Scaler

3.4.2 Encoding data: One-Hot encoding

สำหรับบางเทคนิคที่ Model ไม่รองรับ Categorical Data จะต้องทำการแปลงข้อมูลเป็น Numeric Data ก่อนการนำไปใช้ ผ่านการทำ **One-Hot encoding** ซึ่งคือ การเปลี่ยน Categorical Data ในที่นี้ คือ คอลัมน์ “ภาค” ให้อยู่ในรูปแบบของ Binary values ที่มีค่า 0 หรือ 1 ตาม Value ของข้อมูล ดังรูป

รูปที่ 34



รูปที่ 3-34 ภาพอธิบายหลักการ One-Hot Encoding

เมื่อเตรียมข้อมูลเสร็จแล้วจะได้ข้อมูลพร้อมใช้ ดังรูป

รูปที่ 35

	รายได้เฉลี่ยต่อครัวเรือน (บาท)_Power2	หนี้สินเพื่อการศึกษาระดับปริญญาตรีต่อครัวเรือน (บาท)	ภาค_ภาคใต้	หนี้สินเพื่อการศึกษาระดับปริญญาตรีต่อครัวเรือน (บาท)_Power2	รายได้เฉลี่ยต่อครัวเรือน (บาท)	นร.ภาคจนพิเศษ	ภาค_ภาคเหนือ	ภาค_ภาคกลาง	ภาค_ภาคตะวันออกเฉียงเหนือ	หนี้สินเฉลี่ยต่อครัวเรือน (บาท)	งบประมาณด้านการศึกษา (ล้านบาท)
67	2.216433	0.032667	1.0	0.031589	1.853027	0.120157	0.0	0.0	0.0	1.056425	0.506068
1	1.848400	-0.862552	0.0	-0.452759	1.585696	-0.239742	0.0	1.0	0.0	-0.139661	0.018204
45	-0.013299	0.468511	0.0	0.553911	-0.013521	1.389338	0.0	0.0	1.0	1.618643	0.365291
62	-0.570606	1.454073	0.0	2.426876	-0.624293	0.323063	0.0	0.0	1.0	0.343740	-0.644417
35	-0.582762	-0.667416	0.0	-0.414651	-0.638743	0.065551	1.0	0.0	0.0	0.050543	0.334951
...
28	-0.682110	0.562621	0.0	0.691325	-0.759030	-0.176518	1.0	0.0	0.0	-0.568709	-0.062500
5	0.075965	0.575698	0.0	0.711112	0.076439	-0.196957	0.0	1.0	0.0	-0.589977	-0.580704
23	0.158531	-0.875786	0.0	-0.453981	0.158042	-0.293014	0.0	1.0	0.0	-0.864420	-0.722087
14	-0.154158	-0.711008	0.0	-0.426427	-0.159429	-0.175717	0.0	1.0	0.0	-1.048543	0.039442
4	1.494572	-0.540339	0.0	-0.369613	1.316751	-0.172171	0.0	1.0	0.0	0.122337	1.010316

รูปที่ 3-35 ภาพแสดงข้อมูลทั้งหมดหลังการทำ Data preprocessing

3.4.3 Train Test Split

จะทำการแบ่งชุดข้อมูลเป็น 2 กลุ่ม ได้แก่ Training Data สำหรับการสร้าง Model 80% และ Test Data เป็นข้อมูลสำหรับการทดสอบ 20%

3.4.4 Setup() Model

จาก library Pycaret สามารถทำขั้นตอนดังกล่าวได้ผ่านการใช้ฟังก์ชัน setup() โดยมีรายละเอียดและ default ดังนี้

รูปที่ 36

	Description	Value		Description	Value
0	session_id	842	30	Transformation Method	None
1	Target	จำนวนเด็กหลุดออกจากระบบ (คน)	31	PCA	False
2	Original Data	(77, 8)	32	PCA Method	None
3	Missing Values	False	33	PCA Components	None
4	Numeric Features	5	34	Ignore Low Variance	False
5	Categorical Features	1	35	Combine Rare Levels	False
6	Ordinal Features	False	36	Rare Level Threshold	None
7	High Cardinality Features	False	37	Numeric Binning	False
8	High Cardinality Method	None	38	Remove Outliers	False
9	Transformed Train Set	(61, 11)	39	Outliers Threshold	None
10	Transformed Test Set	(16, 11)	40	Remove Multicollinearity	False
11	Shuffle Train-Test	True	41	Multicollinearity Threshold	None
12	Stratify Train-Test	False	42	Remove Perfect Collinearity	True
13	Fold Generator	KFold	43	Clustering	False
14	Fold Number	5	44	Clustering Iteration	None
15	CPU Jobs	-1	45	Polynomial Features	True
16	Use GPU	True	46	Polynomial Degree	2
17	Log Experiment	False	47	Trigonometry Features	False
18	Experiment Name	reg-default-name	48	Polynomial Threshold	0.100000
19	USI	4539	49	Group Features	False
20	Imputation Type	simple	50	Feature Selection	True
21	Iterative Imputation Iteration	None	51	Feature Selection Method	classic
22	Numeric Imputer	mean	52	Features Selection Threshold	0.800000
23	Iterative Imputation Numeric Model	None	53	Feature Interaction	False
24	Categorical Imputer	constant	54	Feature Ratio	False
25	Iterative Imputation Categorical Model	None	55	Interaction Threshold	None
26	Unknown Categoricals Handling	least_frequent	56	Transform Target	False
27	Normalize	True	57	Transform Target Method	box-cox
28	Normalize Method	robust			
29	Transformation	False			

รูปที่ 3-36 ภาพแสดงข้อมูลรายละเอียดการ Setup() Model

3.4.5 Compare Model

เมื่อทำการ Setup Model เสร็จแล้ว ขั้นตอนถัดไปจะเป็นในส่วนของการสร้าง Model ผ่านฟังก์ชัน Compare Model

เนื่องจากยังไม่สามารถหาข้อสรุปได้ว่าควรสร้าง Model ด้วยเทคนิคไหน ฟังก์ชัน Compare Model จึงเป็นฟังก์ชันที่มีประโยชน์มากในการช่วยเปรียบเทียบผลลัพธ์ของแต่ละ Model ตามข้อมูลที่ Setup ไว้

ข้างต้น และจะทำการวัดประสิทธิภาพของ Model โดยการทำ K-Fold Cross Validation ซึ่งกำหนด K-Fold = 5

Model 1 Model ทำนายจำนวนนักเรียนที่หลุดออกจากระบบการศึกษา

รูปที่ 37

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
gbr	Gradient Boosting Regressor	47.1765	4894.9588	64.2060	0.4908	1.5306	3.4966	0.0640
et	Extra Trees Regressor	46.2765	5345.6864	67.0147	0.4351	1.5767	4.5094	0.5980
ada	AdaBoost Regressor	49.0647	5563.2862	68.1963	0.4040	1.7007	4.6646	0.1480
dt	Decision Tree Regressor	47.9423	5608.9731	71.2442	0.3832	1.5690	2.2475	0.0140
rf	Random Forest Regressor	54.5842	6451.6342	74.8780	0.2924	1.7973	5.3201	0.6000
lightgbm	Light Gradient Boosting Machine	60.8836	7150.3663	83.3800	0.1241	1.9113	5.9806	0.1400
huber	Huber Regressor	52.4235	7951.3763	80.6996	0.0742	1.6184	6.4892	0.0360
dummy	Dummy Regressor	67.2287	8611.2512	91.3808	-0.0446	2.0983	8.0663	0.0060
knn	K Neighbors Regressor	65.0323	9577.0818	93.3194	-0.0802	1.9577	5.4515	0.5880
lr	Linear Regression	71.8032	36266.1435	130.4918	-3.2370	1.7705	12.7835	0.0100
lasso	Lasso Regression	76.1276	59112.1781	152.4981	-5.8815	1.8139	14.9931	0.0120
lar	Least Angle Regression	91.9134	117840.0797	201.7627	-12.7239	1.7961	19.7486	0.0120
par	Passive Aggressive Regressor	92.1785	123119.5494	212.7622	-13.2822	1.5940	16.9731	0.0140
ridge	Ridge Regression	91.8990	126076.1538	203.7760	-13.6448	1.8058	19.6023	0.0080
br	Bayesian Ridge	115.4222	157359.9314	246.1281	-17.3489	2.2519	24.2900	0.0100
llar	Lasso Least Angle Regression	119.8014	243176.0422	277.3578	-27.3043	1.9878	26.1113	0.0100
en	Elastic Net	132.4474	310220.8156	310.0119	-35.0797	2.0790	29.3984	0.0100
omp	Orthogonal Matching Pursuit	185.8727	798708.5204	471.0188	-91.9587	2.2340	45.3048	0.0080

รูปที่ 3-37 ภาพแสดงผลลัพธ์การประเมิน Model 1 ในแต่ละเทคนิค

เพื่อทำนายจำนวนเด็กหลุดออกจากระบบการศึกษา

จากรูปที่ 3-28 ทำให้ทราบว่า Model ที่ดีที่สุดสำหรับการทำนายจำนวนเด็กหลุดออกจากระบบการศึกษา คือ et (Extra Trees Regressor)

Extra Trees เป็น Algorithm ที่ใช้ได้ทั้งโจทย์ Classification และ Regression สำหรับโจทย์นี้จะเลือกใช้ Extra Trees Regressor ในการทำนายจำนวนเด็กที่หลุดออกจากระบบการศึกษาโดย Extra Trees Regressor จะทำงานโดยการสุ่มซับเซตของ Feature และ Training Decision tree กับข้อมูลนั้น หลังจากนั้น Tree จะถูกตัดแต่งเพื่อให้ Feature ที่สำคัญที่สุดสำหรับการทำนายเท่านั้น

Model 2 Model ทำนายจำนวนนักเรียนยากจนพิเศษ

รูปที่ 38

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
rf	Random Forest Regressor	6009.0297	64013811.0620	7787.9901	0.8074	0.7177	0.9110	0.2980
et	Extra Trees Regressor	5404.1700	67640175.2738	8183.0125	0.7934	0.5718	0.6182	0.3420
gbr	Gradient Boosting Regressor	5762.0635	70704132.8871	8105.7668	0.7879	0.6168	0.6339	0.0500
xgboost	Extreme Gradient Boosting	5726.0812	76699973.6050	8189.1632	0.7662	0.7266	0.8593	0.5380
ada	AdaBoost Regressor	6783.3913	86990875.9748	8834.8714	0.7325	0.8177	1.2373	0.0720
knn	K Neighbors Regressor	7466.1713	118022568.5441	10588.9537	0.6551	0.8377	1.1911	0.0140
dt	Decision Tree Regressor	7079.9077	111809586.3385	10382.1836	0.6465	0.6728	0.6634	0.0060
lightgbm	Light Gradient Boosting Machine	10105.8516	180320534.4557	13262.8963	0.4484	1.2031	1.8617	0.1200
omp	Orthogonal Matching Pursuit	10217.4218	201212189.5599	13920.7855	0.3701	1.1287	1.9744	0.0060
dummy	Dummy Regressor	15008.7640	335592735.5010	18250.5032	-0.0157	1.4663	3.7848	0.0040
par	Passive Aggressive Regressor	16908.1065	2488484809.9170	36341.5630	-5.7290	1.1879	4.8877	0.0060
huber	Huber Regressor	14495.2784	2758548865.3335	32881.4622	-6.3597	1.0440	5.5881	0.0220
en	Elastic Net	18584.1507	4591990132.2298	40727.3718	-11.2094	1.2073	7.8587	0.0080
llar	Lasso Least Angle Regression	18397.8268	6641270859.5896	45089.6849	-16.5951	1.2538	8.6811	0.0080
ridge	Ridge Regression	18748.8117	6685629898.4732	45717.3949	-16.7257	1.1246	8.6599	0.0060
lasso	Lasso Regression	18455.9526	6705610157.2645	45276.4952	-16.7651	1.3710	8.7304	0.0080
lr	Linear Regression	18459.6392	6710425741.6127	45289.4693	-16.7778	1.3157	8.7356	0.0060
br	Bayesian Ridge	23260.3112	6813421579.3165	50563.2298	-17.1379	1.5214	10.5423	0.0060
lar	Least Angle Regression	34563.4498	14465298970.5711	75350.0600	-37.7563	1.6160	15.5212	0.0120

รูปที่ 3-38 ภาพแสดงผลการประเมิน Model 2 ในแต่ละเทคนิค

เพื่อทำนาย Model 2 จำนวนนักเรียนยากจน

จากรูปที่ 3-29 ทำให้ทราบว่า Model ที่ดีที่สุดสำหรับการทำนายจำนวนเด็กยากจนพิเศษ คือ rf (Random Forest Regressor)

Random Forest Regressor เป็นเทคนิค Ensemble learning ในการสร้าง Model โดยใช้ Decision Tree แต่เป็นการเพิ่มจำนวนต้นไม้ (Tree) เป็นหลาย ๆ ต้น โดยแต่ละต้นจะได้รับคุณลักษณะ (Feature) และ ข้อมูล (Data) ที่ไม่เหมือนกันทั้งหมด เพื่อให้ได้ต้นไม้ที่มีความหลากหลายและมีความอิสระต่อกันมากขึ้น ทำให้ประสิทธิภาพการทำงานและพยากรณ์สูงขึ้น

3.4.6 Create Model

เมื่อเห็นผลลัพธ์ของแต่ละ Model แล้ว จะทำการเลือก Model ที่มีค่าความแม่นยำสูงที่สุดหรือ ปรากฏเป็นอันดับหนึ่งในตารางมาสร้าง Model ผ่านฟังก์ชัน `create_model('model_name')`

3.4.7 Evaluate Model

สำหรับการประเมิน Model Pycaret มีค่าทางสถิติให้พิจารณาถึง 5 ตัว ได้แก่ MAE, MSE, RMSE, R2, RMSLE และ MAPE โดยเราเลือกพิจารณา MAE, MAPE และ R2 เนื่องจากชุดข้อมูลมีค่า Outlier ค่อนข้างสูง

MAE หรือ Mean Absolute Error เป็นการหาค่าเฉลี่ยของ Absolute ค่า Error (prediction-actual) โดยหากค่าเข้าใกล้ 0 แปลว่า Model ทำนายได้ดี

รูปที่ 39

$$MAE = \frac{1}{n} \sum_{i=1}^n |Error|$$

รูปที่ 3-39 ภาพสูตรการคำนวณ MAE

MAPE หรือ Mean Absolute Percentage Error คือ ค่า MAE แต่ทำให้อยู่ในรูปของ Percent (%) เช่นเดียวกับ MAE หากค่าเข้าใกล้ 0 แปลว่า Model ทำนายได้ดี

รูปที่ 40

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

รูปที่ 3-40 ภาพสูตรการคำนวณ MAPE

R2 หรือ R-Squared เป็นค่าที่แสดงอิทธิพลของตัวแปรอิสระทั้งหมดในสมการที่มีต่อตัวแปรตาม หรือ ตัวแปรอิสระทั้งหมดที่อยู่ในสมการสามารถอธิบายการผันแปรตัวแปรตามได้มากน้อยเพียงใด โดยมีค่าอยู่ระหว่าง 0-1 ค่ายิ่งเข้าใกล้ 1

รูปที่ 41

$$R^2 = 1 - \left(\frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \right)$$

รูปที่ 3-41 ภาพสูตรการคำนวณ R2

3.4.7 Predict Model

เมื่อได้ Model แล้วให้นำมาใช้กับชุดข้อมูล Test set เพื่อดูผลลัพธ์ของ Model ผ่านฟังก์ชัน `predict_model('model_name')`

3.4.8 Feature Important

เป็นคะแนนที่ให้แต่ละ Feature ว่ามีความสำคัญต่อการทำนายแค่ไหน โดยเราจะใช้ Library ที่ชื่อว่า SHAP Value ในการอธิบายว่าแต่ละตัวแปรส่งผลกับผลลัพธ์อย่างไร

3.5. การนำแบบจำลองไปใช้งาน (Model Deployment)

เมื่อได้ Model แล้ว จะทำให้ทราบถึงความสัมพันธ์ของตัวแปรต้นและตัวแปรตามที่มีผลต่อการทำนาย ในขั้นตอนถัดไป นอกจากนี้ยังทราบถึง Feature Importance ว่ามี Feature ใหนบ้างที่มีความสำคัญและมีผลต่อการทำนายของ Model มากที่สุด ซึ่งจะช่วยให้สามารถกำหนดทิศทางหรือนโยบายเพื่อแก้ไขปัญหาเด็กหลุดออกจากระบบการศึกษาได้ต่อไป

บทที่ 4

ความเชื่อมโยงกับวิชาต่าง ๆ ในโมดูล

4.1 วข.310 การสำรวจและการเตรียมข้อมูล

ทำการเตรียมข้อมูลต่าง ๆ โดยใช้เทคนิค Web Scaping ในการรวบรวมข่าวที่เกี่ยวข้องกับเด็กหลุดออกจากระบบการศึกษาให้ออกมาเป็นรูปแบบสกุลไฟล์ csv, xlsx และได้ทำการสรุปผลออกมาเป็นตารางที่แสดงถึง หัวข้อ แหล่งที่มาข่าว เทคนิคที่ใช้ และจำนวนข่าว ได้ดังนี้

ตารางที่ 4-1 ผลลัพธ์ต่าง ๆ ที่ได้จากการทำ Web Scaping

ลำดับ	หัวข้อและแหล่งที่มา	Code GitHub	เทคนิค	จำนวนข่าว
1	ข้อมูลข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์ The Standard	https://github.com/LUNA-eugene/DSI310_group2/blob/main/Analyze_News_1_Process/01_scraping.ipynb	1. Beautiful Soup	21
2	ข้อมูลข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์ Thairath	https://github.com/LUNA-eugene/DSI310_group2/blob/main/News_Thairath_week4/scraping%20thairath%20week4.ipynb	1. Beautiful Soup 2. Selenium	250
3	ข้อมูลข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษาจากเว็บไซต์EEF (กองทุนเพื่อความเสมอภาค)	https://github.com/LUNA-eugene/DSI310_group2/blob/main/Analyze_News_2_Process/scraping_2.ipynb	1. Beautiful Soup	14
รวม				285

รูปที่ 42

Out[46]:	title	content	date
0	พบเด็ก กศน.ค้างท่อหลุดระบบนับแสนคน "วัลลพ" สั...	น.ส.ตรีนุช เทียนทอง รมว.ศึกษาธิการ ได้มีนโยบาย...	2022-08-11
1	สิ้นปีศึกษา 66 นร.ออกกลางคันเป็นศูนย์ "ตรีนุช"...	เปิดภาคเรียนที่ 2/2565 นี้ ศธ.มุ่งเน้นให้ครูกร...	2022-02-11
2	ความยากจนข้ามรุ่น การศึกษาเหลื่อมล้ำ	"ความยากจนข้ามรุ่น" และ "ความเหลื่อมล้ำทางการศ...	2022-01-11
3	ร่วม MOU เดินหน้า "ระบบหลักประกันความเสมอภาคทา...	ร่วม MOU เดินหน้า "ระบบหลักประกันความเสมอภาคทา...	2022-10-31
4	แนะหยุดวงจรความยากจนข้ามรุ่น ด้วยระบบโอกาสการศ...	แนะหยุดวงจรความยากจนข้ามรุ่น ด้วยการพัฒนาระบบ...	2022-10-21
...
250	หรรษา ศก.เข้าเต็ม กศศ.รูกช่วยเด็กยากจน เสีย...	หรรษา ศก.เข้า กศศ.รูกช่วยเด็กยากจน เสีย...	2020-05-03
251	ปูยาติดโควิด-19 จากช็อกโกโด้ ผวาระบาดในไทย	ไทยพบผู้ติดเชื้อไวรัสโควิด-19 เพิ่มขึ้นอีก 3 ร...	2020-02-26
252	"หรรษา" คุย "สี จิ้นผิง" สุวัธน์ มั่นใจจีนชน...	ประธานาธิบดีสี จิ้นผิง ของจีน ยกย่องไทยศ...	2020-07-02
253	ร.ร.ตั้ง ค่อยอดเงินอุดหนุนรายหัว บริจาคให้ กศศ...	สาคัดเกษร อินเดอร์ มุกเบิกค่อยอดเงินอุดหนุนรา...	2020-03-02
254	กศศ.จับมือ กยศ.บูรณาการ ลดปัญหาให้ทุนศึกษา...	กศศ.จับมือ กยศ.บูรณาการฐานข้อมูล ลดปัญหาให้ทุน...	2020-01-29

255 rows x 3 columns

รูปที่ 4-1 ภาพตัวอย่างข้อมูลที่ได้จากการ Web Scraping

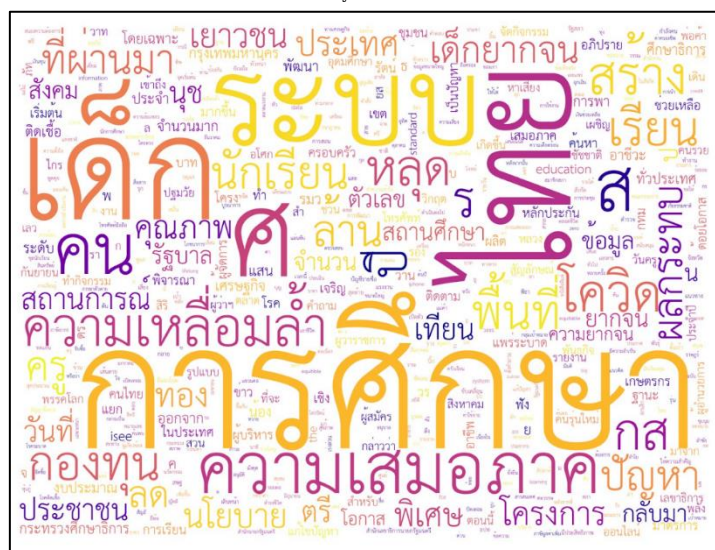
หลังจากทำการ Web Scraping จากทั้ง 3 แหล่งข่าวที่เกี่ยวข้องกับเด็กที่หลุดออกจากระบบการศึกษา แล้วทำการรวมไฟล์ที่ได้จากแหล่งข่าว ให้อยู่ในไฟล์เดียวโดยใช้สกุลไฟล์ xlsx เพื่อนำข้อมูลมาทำ Word Cloud เพื่อดู Keywords ที่เกี่ยวข้องกับปัญหาของเด็กที่หลุดออกจากระบบการศึกษาต่อไป

4.2. วช.311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล

4.2.1 WordCloud

เพื่อหาปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษา โดยเริ่มแรกทำการพิจารณาจากข่าวที่เกี่ยวข้องกับเด็กหลุด รวบรวมข่าวผ่านเทคนิค Web Scraping และนำไฟล์ที่รวมข่าวจากแหล่งข่าวทั้งหมดมาทำ WordCloud เพื่อหาปัจจัยที่ทำให้เด็กหลุดออกจากระบบ โดยแสดงผลออกมาเป็นรูปภาพ ดังนี้

รูปที่ 43



รูปที่ 4-2 ภาพแสดงตัวอย่างของการแสดงผลข้อมูลในรูปแบบ Visualization โดยการนำ WordCloud

จากรูปที่ 4-2 จะเห็นว่า Keywords ที่ถูกกล่าวถึงบ่อย ๆ และสามารถเชื่อมโยงเป็นปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษาได้เช่น ยากจน เด็กยากจน ความยากจน ความเหลื่อมล้ำ ฐานะ เป็นต้น ทำให้สามารถตั้งสมมติฐานได้ว่า ปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุด คือ ความยากจน

Code WordCloud: [https://github.com/LUNA-](https://github.com/LUNA-eugene/DSI314_coding/blob/main/Proccessing%20process/WordCloud.ipynb)

[eugene/DSI314_coding/blob/main/Proccessing%20process/WordCloud.ipynb](https://github.com/LUNA-eugene/DSI314_coding/blob/main/Proccessing%20process/WordCloud.ipynb)

4.2.2 จำแนกเด็กยากจน และคาดว่าจะไม่ยากจนจากข้อมูลเด็กหลุด

ต่อมาจะทำการวิเคราะห์ข้อมูลจากชุดข้อมูลเด็กหลุดออกจากระบบการศึกษา ผ่านการเขียนเงื่อนไขเพื่อจำแนกนักเรียนที่มีสภาวะยากจน และคาดว่าจะไม่ยากจน เพื่อพิสูจน์สมมติฐานที่ว่าปัจจัยที่ส่งผลต่อเด็กหลุดออกจากระบบมากที่สุด คือ ความยากจน และจากการเขียนโปรแกรมนี้ทำให้ทราบถึงข้อมูลเกี่ยวกับจำนวนนักเรียนที่มีสภาวะยากจนจำนวน 2,018 คน และคาดว่าจะไม่ยากจน 3,280 คน

รูปที่ 44

สภาวะความยากจน	
ข้อมูลไม่เพียงพอจึงคาดว่าจะไม่ยากจน	3280
ยากจน	2018

รูปที่ 4-3 ภาพแสดงจำนวนของสภาวะความยากจนของนักเรียน

ซึ่งหากพิจารณาไปถึงความคิดเห็นเพิ่มเติมจะเห็นได้ว่าเป็นข้อมูลที่ไม่ได้กล่าวถึงสาเหตุเพิ่มเติมมากถึง 1,958 คน จึงคาดการณ์ได้นักเรียนที่หลุดออกจากระบบการศึกษาส่วนใหญ่มีสภาวะความยากจน

รูปที่ 45

สภาวะความยากจน	ความเห็นเพิ่มเติม	ความเห็นเพิ่มเติม
ข้อมูลไม่เพียงพอจึง คาดว่าจะไม่ยากจน	-	1958
	ไม่มี	28
	จบการศึกษามัธยมศึกษาตอนต้นเมื่อ 8 เมษายน 2564	13
	จบการศึกษามัธยมศึกษาตอนต้นปี 2563 และศึกษาต่อที่วิทยาลัยอาชีวศึกษาสิริวัฒน จ.นครนายก ตั้งแต่ 25 พฤษภาคม 2564	11
	นักเรียนอายุพ้นเกณฑ์การศึกษามัธยมศึกษาตอนต้น และไม่ประสงค์เรียนต่อ (ได้แนะนำให้เข้าศึกษาต่อ กศน.อ.ยังอ แล้ว)	11
...
ยากจน	ไม่ได้เรียนต่อ จบ ป. 6 ปีการศึกษา 2563 ฐานะยากจนมาก	1
	ไม่ได้เรียนต่อ จบปีการศึกษา 2563	1
	ไม่ได้เรียนต่อมัธยม เนื่องจาก ฝนตก ไม่มีทุนทรัพย์ ตอนนี้ นักเรียนออกทะเลหาปลา กับพ่อ	1
	ไม่แก่นกสัตว์หาย อ.ไม่แก่น จ.ปัตตานี	1
	ได้ให้คำแนะนำไปเรียน กศน. เพราะนักเรียนประสงค์จะทำงานช่วยเหลือครอบครัว	1

รูปที่ 4-4 ภาพแสดงจำนวนข้อมูลสรุปผลรวมของสภาวะความยากจน จำแนกตามความคิดเห็นเพิ่มเติม

Code เงื่อนไขการพิจารณาเด็กยากจน/ไม่ยากจน: https://github.com/LUNA-eugene/DSI314_coding/blob/main/Proccessing%20process/Rule-Based%20NLP%20.ipynb

4.2.3 Correlation Matrix

รูปที่ 46



รูปที่ 4-5 Correlation Matrix แสดงความสัมพันธ์ของตัวแปรทั้งหมด

จากนั้นจะดูความสัมพันธ์เชิงเส้นของแต่ละตัวแปรผ่าน Correlation Matrix ทำให้เราทราบว่าตัวแปรที่มีความสัมพันธ์กับจำนวนเด็กหลุดออกจากระบบมากที่สุดคือจำนวนนักเรียนยากจนพิเศษ และตัวแปรที่สัมพันธ์กับจำนวนเด็กยากจนพิเศษมากที่สุดคือภาคตะวันออกเฉียงเหนือ

4.2.4 Model Regression

จากนั้นได้ทำการสร้าง Model Regression จำนวน 2 Model ได้แก่

1. Model ทำนายจำนวนนักเรียนที่หลุดออกจากระบบการศึกษา
2. Model ทำนายจำนวนนักเรียนยากจนพิเศษ

Model 1 ทำนายจำนวนนักเรียนที่หลุดออกจากระบบการศึกษา

สร้าง Model เพื่อทำนายจำนวนนักเรียนที่หลุดออกจากระบบการศึกษา โดย Feature ที่ใช้เป็น Feature ที่เกี่ยวข้องกับความยากจน เช่น จำนวนนักเรียนยากจนพิเศษ รายได้เฉลี่ยต่อครัวเรือน หนี้สินด้านการศึกษเฉลี่ยต่อครัวเรือน เป็นต้น โดยได้ผลลัพธ์ Model ดังนี้

$$R^2 = 0.8042, \text{MAE} = 29.2131 \text{ และ } \text{MAPE} = 1.6182$$

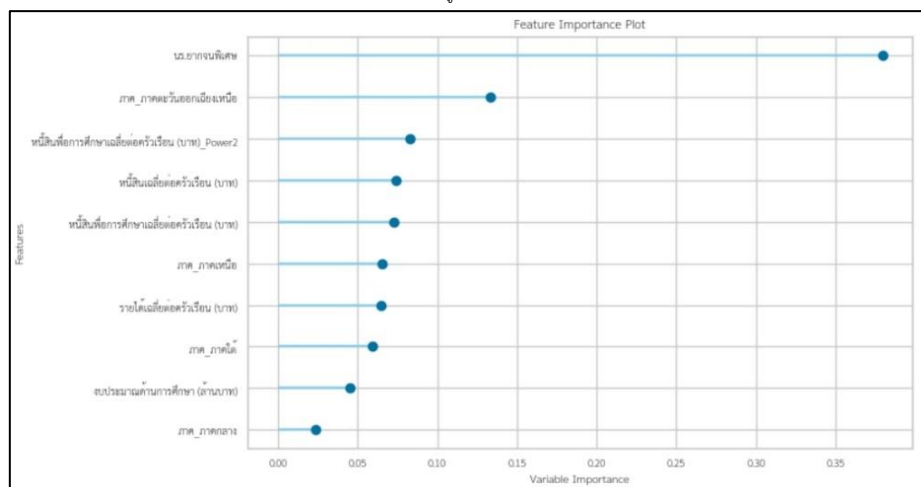
รูปที่ 47

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Extra Trees Regressor	29.2131	1403.4624	37.4628	0.8942	1.6319	1.6182

รูปที่ 4-6 ภาพแสดงผลลัพธ์ความแม่นยำของ Model 1

Feature Important plot

รูปที่ 48

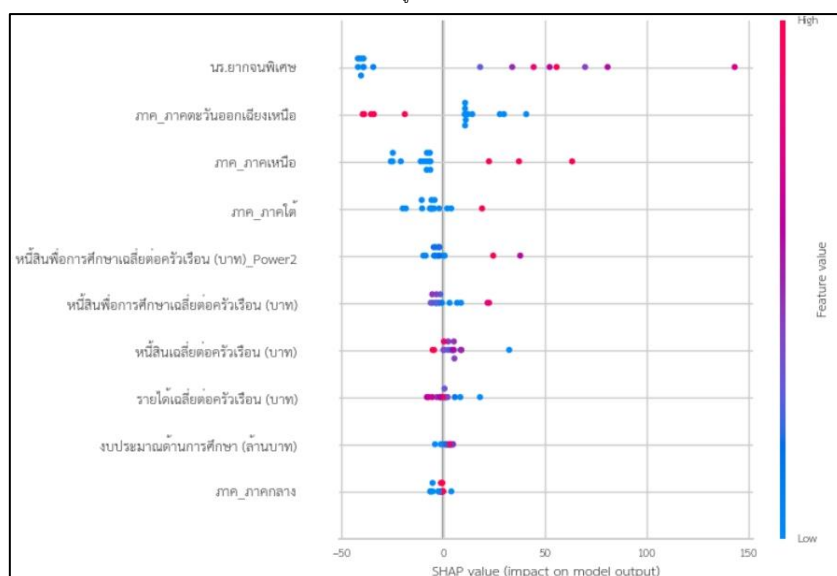


รูปที่ 4-7 ภาพแสดง Feature Important plot ของ Model 1

จากรูปที่ 4-7 ทำให้ทราบ Features ที่มีผลต่อการทำนายจำนวนเด็กหลุดออกจากระบบมากที่สุด 3 อันดับแรก ได้แก่

1. จำนวนนักเรียนยากจนพิเศษ โดยนักเรียนยากจนพิเศษ คือครอบครัวที่มีรายได้เฉลี่ย 1,200 บาทต่อคนต่อเดือน
2. ภาคตะวันออกเฉียงเหนือ
3. ภาคเหนือ

รูปที่ 49



รูปที่ 4-8 ภาพแสดง SHAP Value ของ Model 1

จาก SHAP Value จะสามารถวิเคราะห์ได้ว่าถ้าจำนวนนักเรียนยากจนพิเศษมากจะส่งผลให้มีจำนวนเด็กหลุดออกจากระบบการศึกษามากเช่นกัน แต่เนื่องจากข้อมูลเด็กหลุดออกจากระบบการศึกษาคือเป็นของเดือนกรกฎาคมเพียงแค่เดือนเดียว ทำให้ไม่สามารถสรุปได้ว่าหากเป็นจังหวัดที่อยู่ในภาคตะวันออกเฉียงเหนือจะมีโอกาสที่จำนวนเด็กหลุดออกจากระบบการศึกษาน้อยกว่าอยู่ในภาคอื่น ๆ จึงนำมาสู่ Model ที่ 2 เพื่อทำนายจำนวนนักเรียนยากจนพิเศษต่อไป

Model 2 ทำนายจำนวนนักเรียนยากจนพิเศษ

ได้ผลลัพธ์ Model $R^2 = 0.6747$, $MAE = 4371.1075$ และ $MAPE = 0.7615$

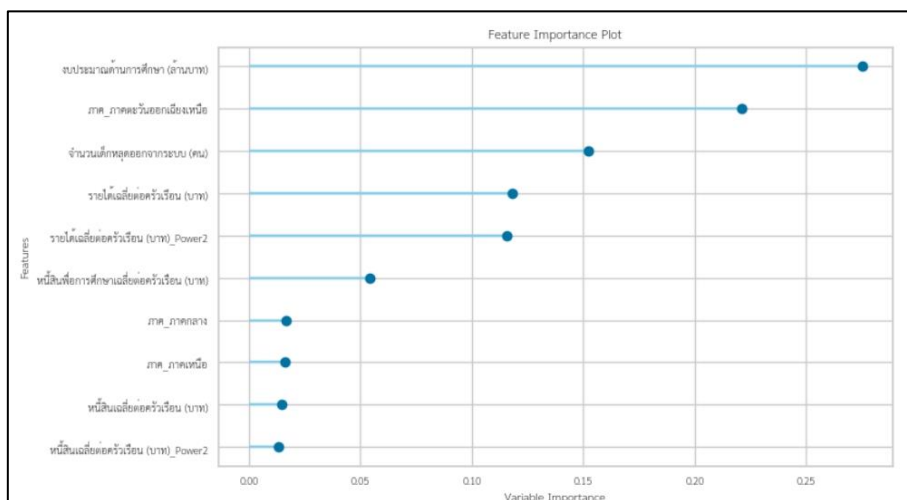
รูปที่ 50

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Random Forest Regressor	4371.1075	41373891.5389	6432.2540	0.6747	0.6915	0.7615

รูปที่ 4-9 ภาพแสดงผลลัพธ์ความแม่นยำของ Model 2

Feature Important plot

รูปที่ 51

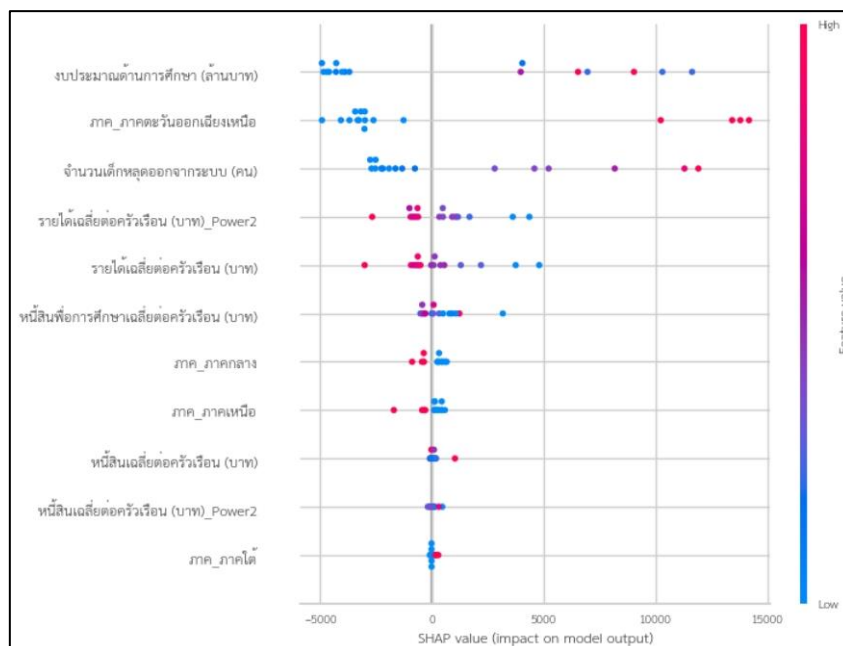


รูปที่ 4-10 ภาพแสดง Feature Important plot ของ Model 2

จากรูปที่ 4-10 ทำให้ทราบถึง Features ที่มีผลต่อการทำนายจำนวนเด็กยากจนพิเศษมากที่สุด 3 อันดับ ได้แก่

1. งบประมาณด้านการศึกษา
2. ภาคตะวันออกเฉียงเหนือ
3. จำนวนเด็กหลุดออกจากระบบการศึกษา

รูปที่ 52



รูปที่ 4-11 ภาพแสดง SHAP Value ของ Model 2

จาก Feature Important plot จะเห็นว่างบประมาณด้านการศึกษามีความสำคัญต่อการทำนายของ Model มากที่สุดก็จริง แต่เมื่อดูที่ SHAP Value จะเห็นถึงความปะปนของ Value มากเกินไปจนทำให้ไม่สามารถสรุปได้ว่า หากมีงบประมาณด้านการศึกษาเยอะจะส่งผลต่อจำนวนเด็กนักเรียนยากจนพิเศษอย่างไร แต่สิ่งที่ SHAP Value บอกได้ มีดังนี้

1. หากเป็นภาคตะวันออกเฉียงเหนือจะมีจำนวนนักเรียนยากจนพิเศษจำนวนมาก
2. หากมีเด็กหลุดออกจากระบบมากจะทำให้มีนักเรียนยากจนพิเศษจำนวนมากตามไปด้วย
3. หากมีรายได้เฉลี่ยต่อครัวเรือนต่ำจะทำให้มีนักเรียนยากจนพิเศษจำนวนมาก

Code Model: [https://github.com/LUNA-](https://github.com/LUNA-eugene/DSI314_coding/blob/main/Processing%20process/Regression%20Model%201%262.ipynb)

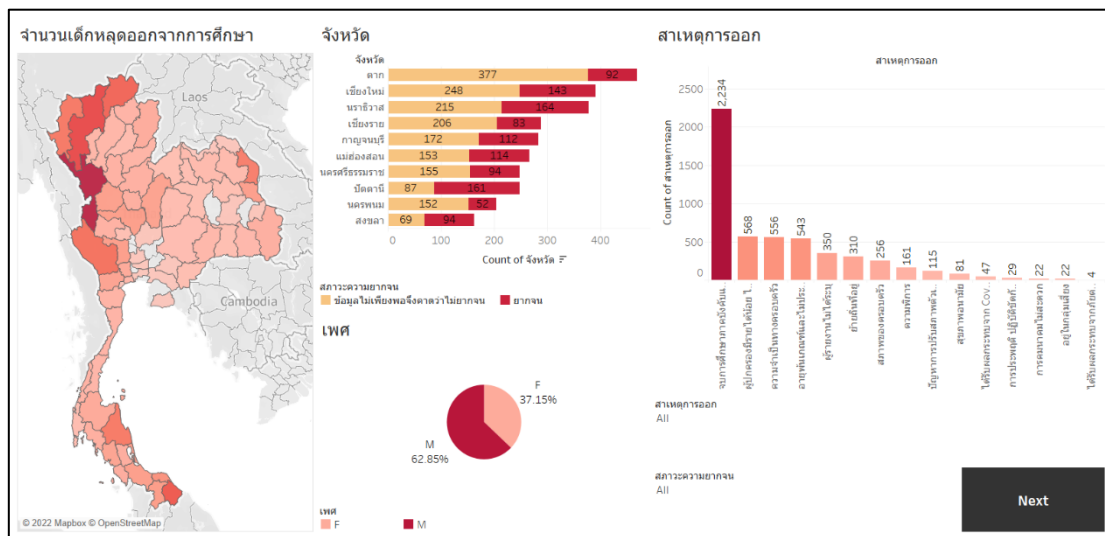
[eugene/DSI314_coding/blob/main/Processing%20process/Regression%20Model%201%262.ipynb](https://github.com/LUNA-eugene/DSI314_coding/blob/main/Processing%20process/Regression%20Model%201%262.ipynb)

4.3 วข.312 ระบบธุรกิจอัจฉริยะ

จากข้อมูลที่ได้ทำการรวบรวมมา เมื่อนำมาสร้าง Dashboard จะได้ผลลัพธ์ ดังนี้

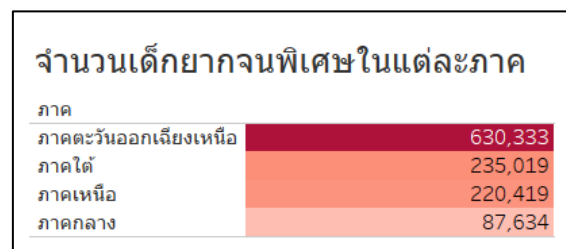
ในเดือนกรกฎาคม ปี 2565 มีเด็กไทยหลุดออกจากการศึกษาทั้งสิ้น 5,298 ราย เป็นเพศชายจำนวน 3,330 ราย หรือคิดเป็น 62.85% และเป็นเพศหญิงจำนวน 1,968 ราย หรือคิดเป็น 37.15% ส่วนใหญ่อาศัยอยู่ในบริเวณภาคเหนือและมีสาเหตุสำคัญมาจากปัญหาความยากจน

รูปที่ 53



รูปที่ 4-12 ภาพ Dashboard แสดงภาพรวมของเด็กหลุดออกจากระบบการศึกษา
เดือนกรกฎาคม 2565

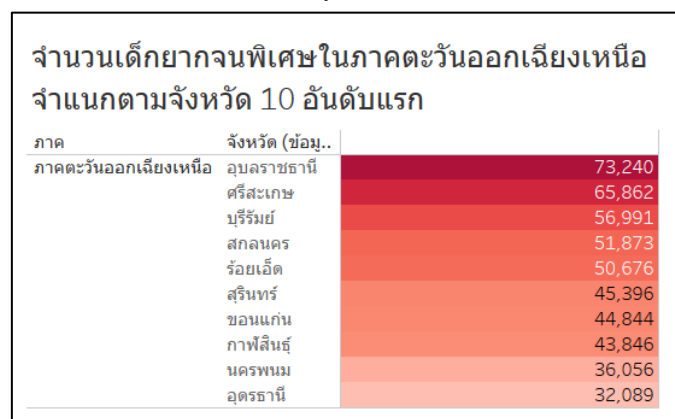
รูปที่ 54



รูปที่ 4-13 ภาพแสดงตารางจำนวนเด็กยากจนพิเศษในแต่ละภูมิภาค

จากรูป และข้อมูลจากการทำ Model ทำให้ทราบว่าหากเป็นจังหวัดที่อยู่ในภาคตะวันออกเฉียงเหนือจะมีโอกาสที่เด็กยากจนพิเศษจะมาก

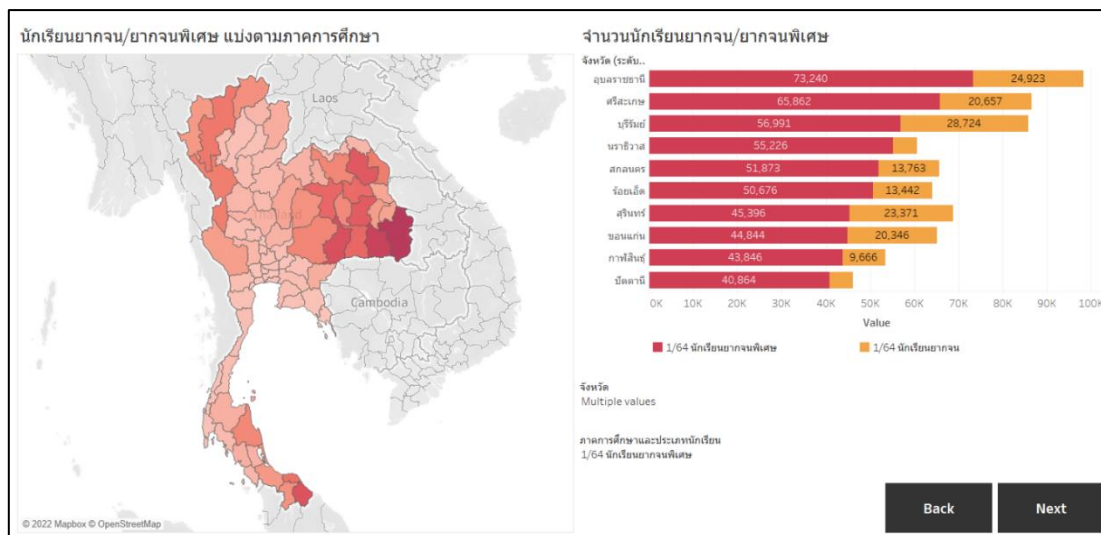
รูปที่ 55



รูปที่ 4-14 ภาพแสดงตารางจำนวนเด็กยากจนพิเศษในแต่ละจังหวัดของภาคตะวันออกเฉียงเหนือ

เมื่อพิจารณาแก่นักเรียนที่มีภาวะยากจนและยากจนพิเศษจะเห็นว่านักเรียนส่วนใหญ่จะอาศัยอยู่ในบริเวณภาคตะวันออกเฉียงเหนือเป็นส่วนใหญ่ เมื่อสนใจเฉพาะนักเรียนยากจนพิเศษ ในปีการศึกษา 2564 ภาคเรียนที่ 1 จังหวัดที่มีจำนวนนักเรียนยากจนพิเศษมากที่สุด ได้แก่ อุบลราชธานี จำนวน 73,240 ราย

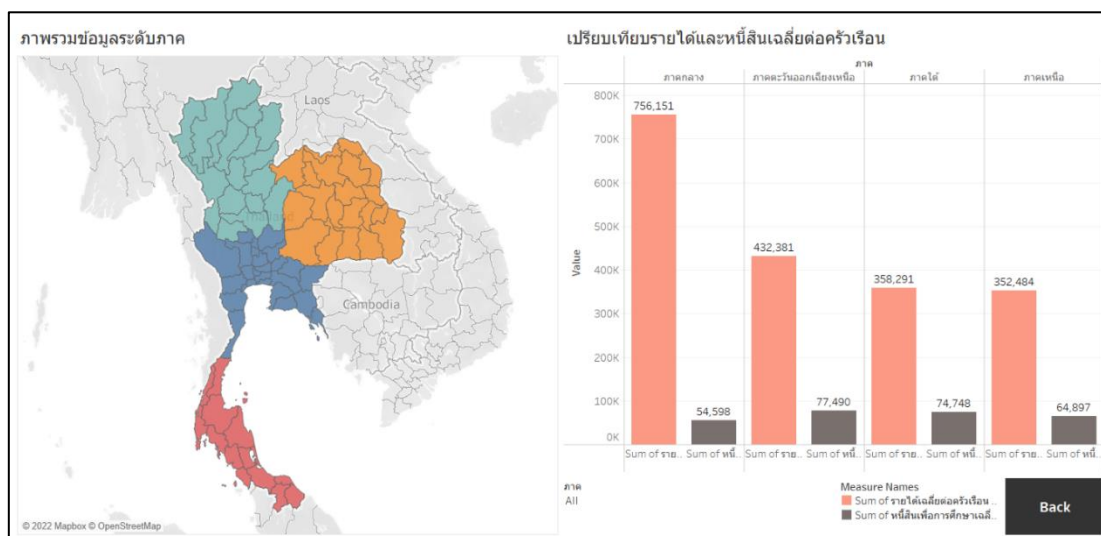
รูปที่ 56



รูปที่ 4-15 ภาพ Dashboard แสดงจำนวนนักเรียนยากจนและนักเรียนยากจนพิเศษ

เมื่อดูภาพรวมในระดับภูมิภาคจะเห็นว่าภูมิภาคที่มีหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือนมากที่สุด คือ ภาคตะวันออกเฉียงเหนือ รองลงมา คือ ภาคใต้และภาคเหนือ ซึ่งมีหนี้สินเฉลี่ยต่อครัวเรือนใกล้เคียงกับภาคตะวันออกเฉียงเหนือมากที่สุดแต่กลับมีรายได้เฉลี่ยต่อครัวเรือนน้อยกว่าถึง 18% และอันดับสุดท้าย คือ ภาคกลาง ซึ่งมีรายได้เฉลี่ยครัวเรือนมากที่สุด ในขณะที่เดียวกันก็มีหนี้สินเพื่อการศึกษาเฉลี่ยต่อครัวเรือนน้อยที่สุดด้วยเช่นกัน

รูปที่ 57



รูปที่ 4-16 ภาพ Dashboard แสดงภาพรวมข้อมูลเปรียบเทียบรายได้และหนี้สินโดยเฉลี่ยในระดับภูมิภาค

ตัวอย่าง Dashboard ทั้งหมด:

https://public.tableau.com/views/Project_16710395778420/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

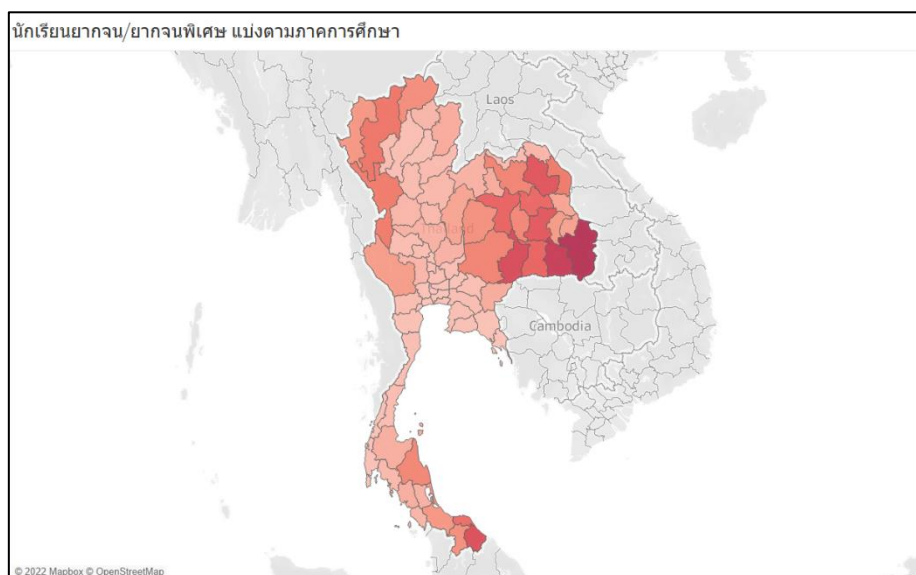
4.4. วช.313 การวิเคราะห์การตลาด

รัฐบาลสามารถนำการวิเคราะห์ข้อมูลเกี่ยวกับเด็กหลุดออกจากระบบการศึกษา เด็กยากจน/ยากจนพิเศษ ไปจนถึงปัจจัยในด้านต่าง ๆ ที่เกี่ยวข้อง มาใช้ในการออกแบบและพัฒนานโยบายเพื่อแก้ปัญหาเด็กหลุดออกจากระบบการศึกษาได้อย่างมีประสิทธิภาพ อีกทั้ง เทศบาลในแต่ละพื้นที่ยังสามารถนำข้อมูลที่ผ่านการวิเคราะห์หรือ Dashboard ไปใช้ในการออกนโยบายหรือแนวทางการแก้ไขปัญหาให้มีความเฉพาะเจาะจงในแต่ละพื้นที่และสามารถแก้ไขปัญหาได้อย่างตรงจุดมากที่สุด

จากการวิเคราะห์ข้างต้นทำให้ทราบว่าปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุด คือ “ความยากจน” และทราบว่าหากจำนวนนักเรียนยากจนพิเศษมีจำนวนมากขึ้นจะทำให้จำนวนเด็กหลุดออกจากระบบการศึกษาเพิ่มขึ้นตามไปด้วย ดังนั้น หากต้องการลดจำนวนเด็กหลุดออกจากระบบก็จำเป็นต้องลดจำนวนเด็กยากจนพิเศษด้วย

จากปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุดคือความยากจน ดังนั้น แนวทางการออกนโยบายการลดจำนวนเด็กหลุดออกจากระบบที่ทางคณะผู้จัดทำเสนอคือการจัดสรรทุนการศึกษาให้เด็กที่หลุดออกจากระบบการศึกษา และแนวทางการป้องกันคือให้ทุนการศึกษากับเด็กยากจนพิเศษ โดยเรามีการตั้งคำถามตามหลัก SMART คือ จากการวิเคราะห์ข้อมูลข้างต้น จะมีวิธีการจัดสรรทุนการศึกษาในระยะเวลา 1 ปี อย่างไร ให้จำนวนเด็กหลุดออกจากระบบการศึกษาลดลง 10% เมื่อเทียบกับปีที่ผ่านมา

รูปที่ 58

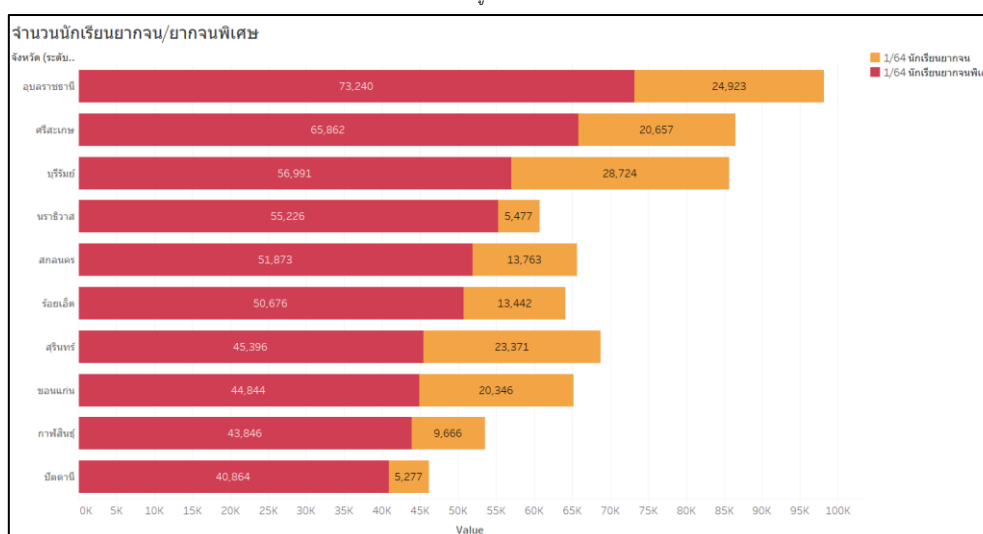


รูปที่ 4-17 ภาพแผนที่แสดงความหนาแน่นของจำนวนเด็กยากจนพิเศษ

สำหรับการจัดสรรทุนการศึกษาจะดำเนินการตามวัตถุประสงค์หลัก 2 ประการ คือ

1. เพื่อลดจำนวนเด็กหลุดออกจากระบบการศึกษาในปัจจุบัน โดยจะพิจารณาการให้ทุนการศึกษากับเด็กที่หลุดออกจากระบบการศึกษา ผ่านการจำแนกเด็กที่มีภาวะยากจนและคาดว่าจะไม่ยากจนที่ได้ดำเนินการในชุดข้อมูลเด็กหลุดออกจากระบบการศึกษา ก.ค. 2565
2. แนวทางป้องกันเด็กหลุดออกจากระบบการศึกษา จากการทราบว่า หากต้องการลดเด็กหลุดออกจากระบบก็จำเป็นที่จะต้องลดจำนวนนักเรียนยากจนพิเศษด้วยเช่นเดียวกัน จึงนำมาสู่การพิจารณาการจัดสรรทุนโดยดูจากจำนวนนักเรียนยากจนพิเศษเป็นหลัก เพื่อลดความเสี่ยงที่เด็กจะหลุดออกจากระบบการศึกษา ดังนั้น จังหวัดที่จะได้รับการพิจารณาให้ทุน 10 อันดับแรก ได้แก่

รูปที่ 59



รูปที่ 4-18 ภาพแสดงจำนวนเด็กยากจนและยากจนพิเศษ 10 อันดับแรก

ตารางที่ 4-2 ตารางแสดงจังหวัดที่ได้รับการพิจารณาการให้ทุน 10 อันดับแรก

ลำดับ	ภาค	จังหวัด	จำนวนนักเรียนยากจนพิเศษ (คน)
1	ภาคตะวันออกเฉียงเหนือ	อุบลราชธานี	72,133
2	ภาคตะวันออกเฉียงเหนือ	ศรีสะเกษ	64,554
3	ภาคตะวันออกเฉียงเหนือ	บุรีรัมย์	56,468
4	ภาคใต้	นราธิวาส	53,859
5	ภาคตะวันออกเฉียงเหนือ	สกลนคร	51,137
6	ภาคตะวันออกเฉียงเหนือ	ร้อยเอ็ด	49,533
7	ภาคตะวันออกเฉียงเหนือ	สุรินทร์	44,530
8	ภาคตะวันออกเฉียงเหนือ	ขอนแก่น	43,536
9	ภาคตะวันออกเฉียงเหนือ	กาฬสินธุ์	43,272
10	ภาคใต้	ปัตตานี	38,548

ตารางที่ 6

จากนั้นจะทำการวิเคราะห์ข้อมูลของแต่ละจังหวัดเพื่อออกนโยบายหรือแนวทางการแก้ไขปัญหา นอกเหนือจากการให้ทุนการศึกษา โดยจังหวัดที่จะนำมาวิเคราะห์เป็นอันดับแรก คือ จังหวัดอุบลราชธานี สำหรับจังหวัดอุบลราชธานี เด็กยากจนส่วนใหญ่ต้องแบ่งเวลาเพื่อช่วยเหลือผู้ปกครองในการทำงาน ทำให้ไม่สามารถเรียนได้อย่างเต็มที่ ซึ่งส่งผลให้เด็กเหล่านั้นอ่านหนังสือไม่ออกและเขียนหนังสือไม่ได้ ดังนั้น จะต้องเพิ่มเวลาเรียนให้เด็กสามารถเรียนตามบทเรียนได้ทันเพื่อให้เกิดการพัฒนาทักษะหรือองค์ความรู้ต่าง ๆ ที่เหมาะสมกับช่วงวัยนั้น ๆ ได้ พร้อมทั้งมีการวางแผนเพื่อแก้ปัญหาในระยะยาวต่อไป ตัวอย่างนโยบายที่ใช้ในการแก้ปัญหา คือ มอบนโยบายให้หน่วยการศึกษาเพื่อหาเวลาสอนเสริมในช่วงพักเที่ยง และให้อาสาสมัครมาเป็นครูสอนเพิ่มในช่วงวันหยุดให้กับเด็ก ๆ เพื่อแก้ปัญหาเด็กที่อ่านหนังสือไม่ออกและเขียนไม่ได้จากการเรียนไม่ทันเพื่อนในชั้นเรียน

บทที่ 5

บทสรุป

ผลการศึกษาปัญหาและปัจจัยที่ทำให้เด็กไทยหลุดออกจากระบบการศึกษา พบว่าปัจจัยที่ทำให้เด็กหลุดจากระบบการศึกษานั้นมีด้วยกันหลายปัจจัย ยกตัวอย่างเช่น ผู้ปกครองมีรายได้น้อย ไม่พอเพียง, การคมนาคมไม่สะดวก, ย้ายถิ่นที่อยู่ และปัญหาการปรับสภาพตัวเองในสังคม เป็นต้น ทางคณะผู้จัดทำได้ทำการวิเคราะห์เพื่อหาปัจจัยที่ทำให้เด็กหลุดออกจากระบบมากที่สุด ซึ่งสามารถสรุปได้ว่า “ความยากจน” เป็นปัจจัยที่ทำให้เด็กหลุดออกจากระบบการศึกษามากที่สุด ดังนั้น ถ้าหากต้องการลดจำนวนเด็กหลุดออกจากระบบการศึกษา ก็จำเป็นที่จะต้องแก้ไขปัญหารื่องความยากจนก่อนเป็นอันดับแรก

ทางคณะผู้จัดทำจึงได้คิดหาแนวทางในการช่วยแก้ไขปัญหาคเด็กหลุดออกจากระบบการศึกษา เพื่อลดโอกาสที่เด็กจะหลุดจากระบบการศึกษา และในขณะเดียวกันก็เป็นการสร้างโอกาสให้เด็กเติบโตไปได้อย่างมีคุณภาพชีวิตและอนาคตที่ดีต่อไป โดยการพิจารณาการให้ทุนการศึกษาแก่เด็กยากจนพิเศษ ซึ่งจะทำการพิจารณาการให้ทุนแก่จังหวัดที่มีจำนวนเด็กยากจนพิเศษมากที่สุดก่อน เนื่องจากเป็นจังหวัดที่มีความเสี่ยงต่อการที่เด็กจะหลุดจากระบบการศึกษามากที่สุด

จากการวิเคราะห์สามารถสรุปได้ว่า จังหวัดอุบลราชธานี เป็นจังหวัดที่มีปัญหานักเรียนยากจนพิเศษมากที่สุดในประเทศไทย ดังนั้น จึงได้รับการพิจารณาการให้ทุนการศึกษาเป็นอันดับแรก และข้อมูลที่คณะผู้จัดทำได้ทำการวิเคราะห์และสรุปผลมาในรูปแบบของ Dashboard โดยหน่วยงานภาครัฐหรือองค์กรการกุศลต่าง ๆ สามารถนำ Dashboard ของคณะผู้ทำไปเป็นข้อมูลเพื่อใช้ประกอบการพิจารณาการจัดสรรงบประมาณ หรือทุนการศึกษาและออกแบบนโยบายที่มีความเฉพาะเจาะจงและเหมาะสมกับแต่ละจังหวัดได้

บรรณานุกรม

Pycaret. (2022). *Quickstart*. สืบค้นจาก

<https://pycaret.gitbook.io/docs/get-started/quickstart>

Pycaret. (2022). *Welcome to PyCaret*. สืบค้นจาก

<https://pycaret.gitbook.io/docs/>

Regression Tutorial (REG101) - Level Beginner. (2020). สืบค้นจาก

<https://nbviewer.org/github/pycaret/pycaret/blob/master/tutorials/Regression%20Tutorial%20Level%20Beginner%20-%20REG101.ipynb>

Pablo Aznar. (2022). *What is the difference between Extra Trees*

and Random Forest?. สืบค้นจาก <https://quantdare.com/what-is-the-difference-between-extra-trees-and-random-forest/>

PCP. (2020). *MAE, MSE, RMSE — เลือกใช้ยังไงดีมาลองดูที่ความหมาย*. สืบค้นจาก

<https://shorturl.asia/Vrpxb>

PCP. (2020). *Pycaret — ผู้ช่วยยุคใหม่ที่จะทำให้การเลือกใช้ Machine Learning ง่ายขึ้น*. สืบค้นจาก

<https://shorturl.asia/UA7Hg>

สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. (2022). รายงาน

วิเคราะห์สถานการณ์ความยากจนและความเหลื่อมล้ำของประเทศไทย ปี 2564. สืบค้นจาก [ewt_dl_link.php \(nesdc.go.th\)](http://ewt_dl_link.php(nesdc.go.th))

อนรรฆ พิทักษ์ธานิน. (2022). *นวัตกรรมการคนทำงานด้านหน้า ‘อุปสรรค-ปัจจัย’ ต่าง ๆ ที่ทำให้เด็ก*

หลุดจากระบบการศึกษา. สืบค้นจาก <https://shorturl.asia/YNGTi>

ศรธรรม หงษ์พรหม. (2020). *การทำนายระดับความยากจนจากของข้อมูลสำมะโนประชากรด้วยการเรียนรู้ของ*

เครื่อง PREDICTION OF POVERTY LEVEL ON CENSUS DATA USING MACHINE LEARNING.

สืบค้นจาก [gs601130298.pdf \(swu.ac.th\)](https://gs601130298.pdf(swu.ac.th))