รายงานวิจัยฉบับสมบูรณ์

ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยขอนแก่น

Success Factors on Graduation of Under-Graduation Students, Khon Kaen University

คณะผู้วิจัย
นายสุภาพ ไชยยา
นายวิทวัส เหล่ามะลอ

โครงการวิจัยนี้ได้รับทุนสนับสนุนโครงการวิจัยจาก
สำนักบริหารและพัฒนาวิชาการ มหาวิทยาลัยขอนแก่น
ประจำปีงบประมาณ 2565

บทคัดย่อ

การศึกษาครั้งนี้ศึกษาว่ามีปัจจัยอะไรบ้างที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรีของมหาวิทยาลัยขอนแก่นบ้าง ได้แก่ 1) ปัจจัยทางด้านส่วนตัว เช่น เพศ รายได้ของครอบครัว การประกอบอาชีพของผู้ปกครอง 2) ปัจจัยทางด้านการศึกษา เช่น คะแนนรวม เกรดเฉลี่ยทั้ง 2 ภาคในการเรียนในชั้นปีที่ 1 คะแนนที่ผ่านการคัดเลือก และ 3) ปัจจัยในการให้บริการของมหาวิทยาลัยขอนแก่น คือ ค่าเฉลี่ยคะแนนของการประเมินการเรียนการสอบของอาจารย์ประจำวิชาและผลการประเมินรายวิชาที่ลงทะเบียนเรียน โดยนำข้อมูลจากฐานข้อมูลรับเข้าศึกษาและฐานข้อมูลทะเบียนนักศึกษา ที่เข้าศึกษาด้วยวิธีรับตรงประจำปีการศึกษา 2561 จำนวน 1,641 รายการ เมื่อทำความสะอาดข้อมูลโดยลบข้อมูลที่ไม่สมบูรณ์ หรือ ข้อมูลที่มีค่าผิดปกติแล้วทำให้มีข้อมูลสำหรับวิเคราะห์ทั้งสิ้น 1,512 รายการ วิเคราะห์ด้วยวิธีการทางสถิติเชิงอนุมาน (Inferential Statistics Analysis) โดยใช้โปรแกรม Python ด้วยวิธี Logistic Regression และ Decision Tree ผ่าน Google Collaboratory

ผลการศึกษาปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) ด้วย Logistic Regression Model พบว่าฟีเจอร์ที่มีผลต่อการทำนายสูงที่สุด คือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE), เพศ และ ผลคะแนนภาษาอังกฤษ จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา ในขณะที่การพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาที่รับด้วยวิธีรับตรงประจำปีการศึกษา 2561 ด้วย Decision Tree Model พบว่าปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) สูงที่สุด คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE) และ ระดับคะแนนเฉลี่ยการประเมินรายวิชาที่เรียน จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา

คำสำคัญ Logistict Regression, Decision Tree, การรับเข้าศึกษา

กิตติกรรมประกาศ

ทีมงานผู้วิจัยขอขอบคุณผู้บริหารสำนักบริหารและพัฒนาวิชาที่ได้ให้ทุนสำหรับการทำวิจัยในครั้งนี้ และขอขอบคุณ รศ. ดร. วรารัตน์ สงฆ์แป้น รองผู้อำนวยการฝ่ายเทคโนโลยีสารสนเทศ และนวัตกรรมดิจิทัล สำนักบริหารและพัฒนาวิชาการ ที่ให้ความกรุณาในการให้คำปรึกษาในการวิเคราะห์ข้อมูล อา กรรมการผู้ทรงคุณวุฒิงานวิจัยของสำนักบริหารและพัฒนาวิชา อาจารย์ ธนพล ตั้งชูพงศ์ อาจารย์ประจำคณะวิทยาลัยการคอมพิวเตอร์และทีมงานที่ได้ให้ความรู้ในเรื่องของการวิเคราะห์ข้อมูล (Data Science And Application)

สารบัญ

หน้า

บทคัดย่อ ..... *ก*

กิตติกรรมประกาศ ..... *ข*

สารบัญ ..... *ค*

สารบัญตาราง ..... *จ*

สารบัญรูปภาพ ..... *ฉ*

บทที่ 1 บทนำ *1*

1.1 ความสำคัญและที่มา 1

1.2 วัตถุประสงค์ของการวิจัย 1

1.3 ขอบเขตการวิจัย 1

1.4 ประโยชน์ที่คาดว่าจะได้รับ 1

1.5 นิยามคำศัพท์ที่ใช้ในการวิจัย 2

1.6. กรอบแนวคิดในการวิจัย 2

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง *3*

2.1 วิทยาการข้อมูล (Data Science) 3

2.2 เป้าหมายของการทำการวิเคราะห์ข้อมูล 3

2.3 ขั้นตอนการวิเคราะห์ข้อมูล (Data Analytics) 4

2.4 กระบวนการทางวิทยาการข้อมูล (Data Science Process) 5

2.5 ประโยชน์ของการทำวิเคราะห์ข้อมูล 7

2.6 งานวิจัยที่เกี่ยวข้อง 8

บทที่ 3 วิธีดำเนินการศึกษาวิจัย *11*

3.1 ประชากรและกลุ่มตัวอย่าง 11

3.2 เครื่องมือที่ใช้ในการวิจัย 11

บทที่ 4 ผลการศึกษา *12*

4.1 การเตรียมข้อมูล 12

4.2 Import Library ที่จะใช้งานและโหลดข้อมูล 13
4.3 สำรวจข้อมูล 15
4.4 การแก้ปัญหา Multicollinearity 17
4.5 การฝึกโมเดล 19
4.6 ทดสอบความแม่นยำของโมเดล 21
4.7 นำไปใช้พยากรณ์ 25
4.8 ทำนายผลด้วยการใช้ Decision Tree Model 26
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ *30*
5.1 สรุปผลการวิจัย 30
5.2 ประสิทธิภาพของการทำนายผลของโมเดล 31
5.3 อภิปรายผลการวิจัย 32
5.3 ข้อเสนอแนะ 32
เอกสารอ้างอิง *34*
ภาคผนวก *38*
ประวัติผู้วิจัย *39*

สารบัญตาราง

ตารางที่ 1 แสดง Confusion Metrics 6

สารบัญรูปภาพ

รูปภาพที่ 1 แสดงประเภทของการวิเคราะห์ข้อมูลและความสัมพันธ์ระหว่างความยากง่ายและมูลค่าที่เกิดขึ้นในขั้นตอนการวิเคราะห์รูปแบบต่างๆ (Elliott T., 2013).....4
รูปภาพที่ 2 แสดงข้อมูลข้อมูล 5 รายการแรกด้วยคำสั่ง head(5).....14
รูปภาพที่ 3 ผลของการใช้คำสั่ง info() สำหรับสำรวจข้อมูลของ Dataset.....14
รูปภาพที่ 4 แสดงข้อมูลจำนวนผู้สำเร็จการศึกษาแยกตามเพศ.....15
รูปภาพที่ 5 แสดง GPA เฉลี่ยของนักศึกษาแยกตามการสำเร็จการศึกษาและเพศ.....16
รูปภาพที่ 6 แสดงข้อมูลแต่ละฟิลด์ที่มีค่าว่าง.....17
รูปภาพที่ 7 แสดงข้อมูลที่ตัดเอาปัจจัยที่ไม่นำมาวิเคราะห์ออกแล้ว.....18
รูปภาพที่ 8 แสดงความสัมพันธ์ (Correlation) ของข้อมูลในภาพรวม.....18
รูปภาพที่ 9 แสดง GPA2 ที่มีค่าว่าง.....19
รูปภาพที่ 10 แสดงผลการทำนายการสำเร็จการศึกษา.....21
รูปภาพที่ 11 แสดง Confussion Metrics ของ Logistict Regression Model ที่ทำนายครั้งแรก 22
รูปภาพที่ 12 แสดง Accuracy, Precision, Recall และ F1-score ในการทำนายครั้งแรก.....22
รูปภาพที่ 13 แสดงค่า Accuracy, Precision, Recall, F1-Score หลังปรับ Hyperparameter 23
รูปภาพที่ 14 แสดงค่า Confussion Metrics หลังปรับ Hyperparameter.....24
รูปภาพที่ 15 แสดงค่า Accuracy, Precision, Recall, F1-Score หลังปรับค่า K-Folds Cross Validation.....25
รูปภาพที่ 16 แสดงปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษากลุ่มตัวอย่าง.....25
รูปภาพที่ 17 แสดงค่า Accuracy, Precision, Recall, F1-Score ของ Decision Tree Model....26
รูปภาพที่ 18 แสดง Confussion Metrics ของ Decition Tree Model.....27
รูปภาพที่ 19 แสดงผลของการปรับพารามิเตอร์เพื่อเพิ่มประสิทธิภาพของ Decision Tree Model 28
รูปภาพที่ 20 แสดงภาพการทำนายของ Decision Tree Model.....28

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มา

การสำเร็จการศึกษาของนักศึกษาถือว่าเป็นตัวชี้วัดที่สำคัญในการวัดความสำเร็จของการผลิตบัณฑิตของสถาบันการศึกษาต่างๆ ในปัจจุบันมหาวิทยาลัยต่างๆ ยังไม่ได้นำเอาข้อมูลขนาดใหญ่ (Big Data) มาใช้วิเคราะห์เพื่อหาความสัมพันธ์ของปัจจัยต่างๆ ของนักศึกษาว่ามีผลต่อการสำเร็จการศึกษาของนักศึกษาหรือไม่ อย่างไร ซึ่งแตกต่างกับภาคเอกชนที่ได้นำการวิเคราะห์ข้อมูลขนาดใหญ่มาใช้ในการดำเนินธุรกิจให้สามารถแข่งขันในยุคดิจิทัลได้

การศึกษาครั้งนี้จึงต้องการศึกษาว่ามีปัจจัยอะไรบ้างที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรีของมหาวิทยาลัยขอนแก่นบ้าง ได้แก่ 1) ปัจจัยทางด้านส่วนตัว เช่น เพศ รายได้ของครอบครัว การประกอบอาชีพของผู้ปกครอง 2) ปัจจัยทางด้านการศึกษา เช่น ขนาดของโรงเรียนและภาคที่ตั้งของโรงเรียน เกรด ม. ปลาย เกรดในการเรียนในชั้นปีที่ 1 คะแนนที่ผ่านการคัดเลือก และ 3) ปัจจัยในการให้บริการของมหาวิทยาลัยขอนแก่น เช่น การเข้าใช้บริการ Internet การใช้บริการห้องสมุด และการประเมินการเรียนการสอน ฯลฯ

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยขอนแก่น

1.3 ขอบเขตการวิจัย

ศึกษาศึกษาปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรีที่รับเข้าด้วยวิธีรับตรงในปีการศึกษา 2561

1.4 ประโยชน์ที่คาดว่าจะได้รับ

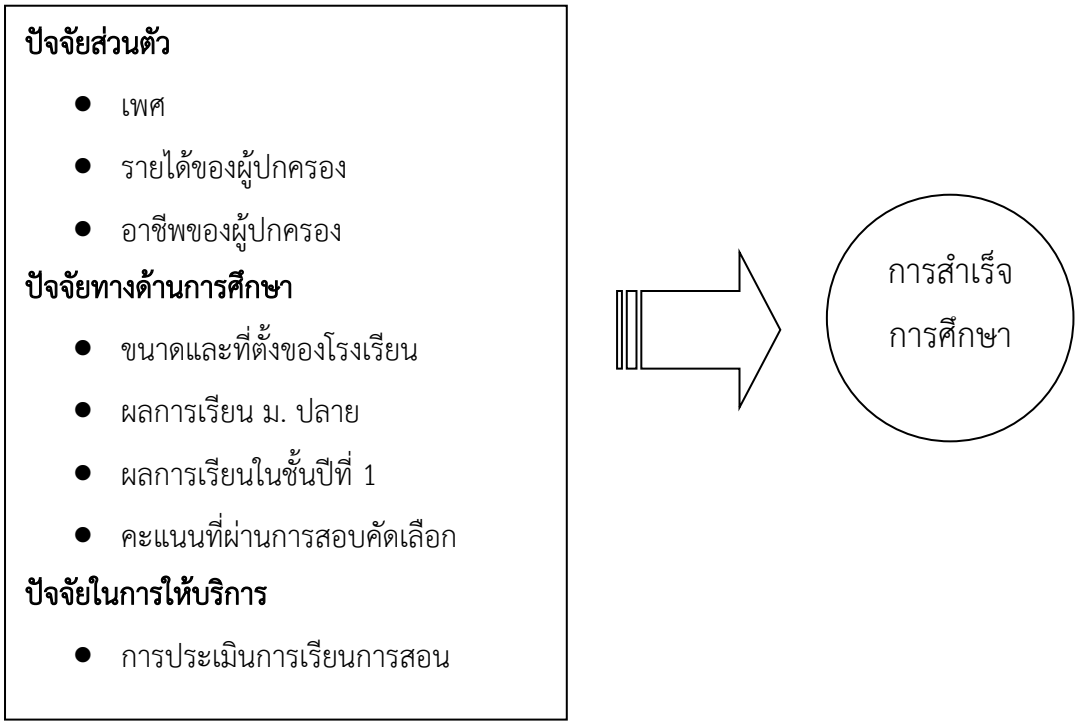
ได้ทราบปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยขอนแก่น

1.5 นิยามคำศัพท์ที่ใช้ในการวิจัย

Logistic Regression หมายถึง โมเดลที่ใช้สำหรับการวิเคราะห์มูลในกลุ่ม Clasification

Decision Tree หมายถึง โมเดลที่ใช้สำหรับการวิเคราะห์มูลในกลุ่ม Clasification

1.6. กรอบแนวคิดในการวิจัย



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยขอนแก่นผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี และ งานวิจัยที่เกี่ยวข้อง ดังนี้

2.1 วิทยาการข้อมูล (Data Science)

วิทยาการข้อมูลเป็นการเรียนรู้เกี่ยวกับสิ่งที่เราสนใจจากข้อมูลโดยอาศัยกระบวนการทางวิทยาการข้อมูล เช่น การค้นหาข้อมูล (Exploration) เพื่อหารูปแบบของข้อมูล การอนุมานข้อมูล (Inference) เพื่อหาข้อสรุปด้วยการใช้สถิติต่างๆ และ การทำนายหรือพยากรณ์ (Prediction) จากข้อมูลที่มีอยู่ด้วย machine learning เป็นต้น (ตั้งชูพงศ์, 2565) การวิเคราะห์ข้อมูลจะทำให้บริษัทได้ข้อมูลเชิงลึกเพื่อนำไปกำหนดแผนดำเนินการทำให้ได้กำไรมากยิ่งขึ้นหรือค้นพบข้อมูลเชิงลึกใหม่ๆ (Subramaniam, A.,2022)

Bahga, A., & Madisetti, V. (2019) กล่าวว่าการวิเคราะห์ข้อมูลเป็นคำที่กว้างที่ครอบคลุมไปถึงขบวนการ เทคโนโลยี กรอบโครงสร้างและอัลกอลิธึมในการสกัดข้อมูลที่มีประโยชน์ ไม่ว่าจะทำการกรอง เปลี่ยนแปลง จัดกลุ่ม เป็นต้น วิทยาการข้อมูลไม่ใช่การสร้างโมเดลที่ซับซ้อนหรือสร้าง Visualization ที่สวยงาม หรือการเขียนโค้ด แต่เป็นการสร้างผลกระทบ (Impact) ให้กับหน่วยงานให้มากที่สุด ไม่ว่าจะเป็นข้อมูลเชิงลึก หรือผลิตภัณฑ์ใหม่ให้กับหน่วยงาน (Joma Tech., 2018)

2.2 เป้าหมายของการทำการวิเคราะห์ข้อมูล

เป้าหมายของการทำการวิเคราะห์ข้อมูล จะทำเพื่อ

1. ทำนายหรือพยากรณ์อะไรสักอย่าง เช่น การทำธุรกรรมนี้เป็นการโกงหรือไม่ วันนี้ฝนจะตกหรือไม่ หรือ ก้อนเนื้อนี้เป็นอันตรายหรือไม่

2. หารูปแบบของข้อมูล เช่น วันที่หนาวที่สุด 10 อันดับในรอบปี หาว่าเพจหน้าไหนที่คนเข้าชมเยอะที่สุด หรือบุคคลสำคัญคนไหนถูกค้นข้อมูลมากที่สุด และ

3. หาความสัมพันธ์ของข้อมูล เช่น หาข่าวที่ใกล้เคียงกัน หาคนไข้ที่มีอาการเหมือนกันในฐานข้อมูล หาข้อมูลสินค้าที่ใกล้เคียงกันที่ลูกค้าสั่งออนไลน์ หารูปที่มีลักษณะเหมือนกัน เป็นต้น

2.3 ขั้นตอนการวิเคราะห์ข้อมูล (Data Analytics)

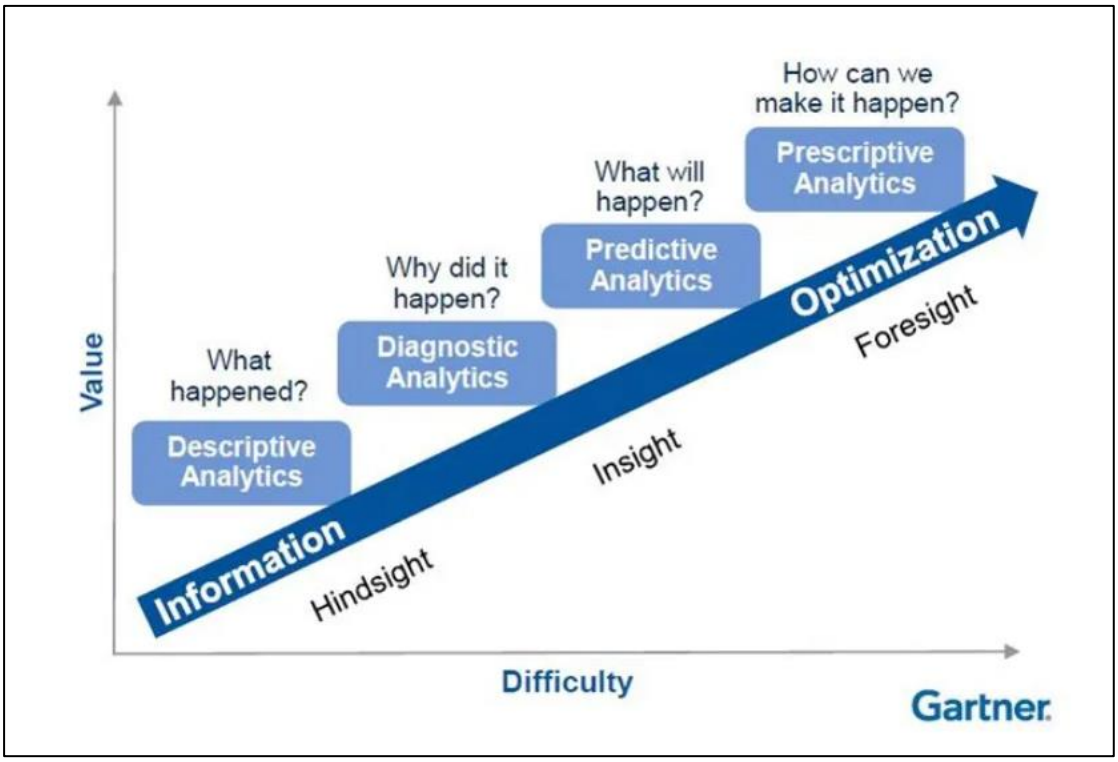
เป็นขั้นตอนที่เปลี่ยนข้อมูลที่มีจำนวนมากให้เป็นข้อมูลเชิงลึกที่มีคุณค่าและนำไปใช้ได้จริง การวิเคราะห์ข้อมูลมีหลายประเภทตามวัตถุประสงค์ของการวิเคราะห์ข้อมูลที่กำหนดในขั้นตอนที่ 1 ว่าต้องการหาข้อมูลเชิงลึกที่เป็นประโยชน์ในระดับไหน การวิเคราะห์ข้อมูลออกเป็น 4 ประเภทดังนี้

1) การวิเคราะห์แบบพื้นฐาน (Descriptive Analytics) จะได้คำตอบว่าเกิดอะไรขึ้น ใช้ข้อมูลในอดีตที่มีเพื่อหาแนวโน้ม หรือรูปแบบของข้อมูล

2) การวิเคราะห์แบบวินิจฉัย (Diagnostic Analytics) ทำไมถึงเป็นเช่นนั้น หาเหตุผลว่าเกิดขึ้นได้เพราะอะไร มีปัจจัยอะไรทำให้เกิดแนวโน้มแบบนั้น ทำไมจึงเกิดเหตุการณ์นั้นได้

3) การวิเคราะห์แบบพยากรณ์ (Predictive Analytics) แล้วจะเกิดอะไรขึ้น ใช้สถิติช่วยทำนายผลลัพธ์ที่จะเกิดขึ้นในอนาคต ถ้าสามารถทำได้ให้ลองพิจารณาว่าเป็นแนวโน้มที่คาดว่าจะเกิดขึ้นต่อเนื่องหรือเกิดซ้ำหรือไม่เพื่อลดความเสี่ยงในการตัดสินใจให้กับองค์กร

4) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive Analytics) จะทำให้เกิดผลลัพธ์ขึ้นได้อีกหรือไม่ ทำได้อย่างไร ด้วยการใช้ AI หรือ Machine Learning ช่วยทำนาย (AIT Share, 2020; Gavin M., 2019; Cote C, 2021; Elliott T., 2013; Bahga, A., & Madisetti, V., 2019)



รูปภาพที่ 1 แสดงประเภทของการวิเคราะห์ข้อมูลและความสัมพันธ์ระหว่างความยากง่ายและมูลค่าที่เกิดขึ้นในขั้นตอนการวิเคราะห์รูปแบบต่างๆ (Elliott T., 2013)

2.4 กระบวนการทางวิทยาการข้อมูล (Data Science Process)

กระบวนการทางวิทยาการข้อมูล มีขั้นตอนดังนี้

1. การตั้งคำถามที่น่าสนใจ เช่น วัตถุประสงค์ในการวิเคราะห์ครั้งนี้คืออะไร ผู้บริหาร/องค์กรอยากได้อะไร ถ้าเรามีข้อมูลอยู่ในมือ สิ่งที่เราสามารถทำได้มีอะไรบ้าง ต้องการพยากร หรือ ประมาณค่าอะไร หรือ ต้องการประมาณรายจ่าย ต้องการพยากรรายได้

2. การเก็บข้อมูล ข้อมูลที่เก็บมานั้น มีการเก็บมาในลักษณะใด ข้อมูลไหนบ้างที่เกี่ยวข้องกับงานหรือไม่ มีข้อพึงระวังทางด้านความเป็นส่วนตัวของผู้ให้ข้อมูล หรือไม่

3. การสำรวจข้อมูล เพื่อใช้การแสดงผลข้อมูลในรูปแบบต่าง ๆ ร่วมกับสถิติพรรณนาในการช่วยศึกษา และทำความเข้าใจคุณลักษณะของข้อมูล มีในส่วนของข้อมูลที่ผิดปกติ หรือไม่ หรือพบปัญหากับข้อมูลที่เก็บมาหรือไม่ ข้อมูลที่สนใจนั้นมีรูบแบบหรือความสัมพันธ์ที่น่าสนใจหรือไม่

4. การสร้างโมเดลของข้อมูล ปัญหาที่สนใจนั้นจัดอยู่ในปัญหาแบบใหน และจะทำการโมเดลข้อมูลที่เราสนใจนี้ โดยใช้ตัวแบบอะไร ทำการฝึกสอนตัวแบบ ทำการตรวจสอบความถูกต้องของตัวแบบ

4.1 รูปแบบจำลองจากการเรียนรู้ของเครื่อง (Machine Learning Models)

สามารถแบ่งรูปแบบจำลองจากการเรียนรู้ของเครื่องได้เป็น 4 ประเภท (AIT Share, 2022a)

1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการเรียนรู้จากตัวอย่าง
2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้จากากรสังเกต
3) การเรียนรู้แบบเสริมกาลัง (Reinforcemen Learning) เป็นการเรียนรู้จากการให้รางวัล หรือ ถ้าทำผิดให้ลงโทษ
4) การเรียนรู้แบบเพื่อสร้าง (Generative Learning) เป็นการเรียนรู้จากการสังเกตและสร้าง

4.2 อัลกอริทึม

อัลกอริทึมคือกระบวนการประมวลผลข้อมูลที่นำเข้ามาในระบบเพื่อใช้สร้างโมเดลการวิเคราะห์ข้อมูล ถ้าเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) จะแบ่งออกเป็น Regression และ Classification ส่วนการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) จะแบ่งออกเป็น Clustering และ Dimensional Reduction (AIT Share, 2022b)

4.3 ตัวชี้วัด (Classification Metrics)

ตัวชี้วัดจะใช้ Confusion Metrics ดังแสดงในตารางที่ 1 (AIT Share, 2022c, Dataike, 2021)

		Actual	
		Positive	Negative
Preditable	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

ตารางที่ 1 แสดง Confusion Metrics

ซึ่งจะมีค่าดังต่อไปนี้

1) Accuracy ค่าความแม่นยำ เหมาะสำหรับ Unbiased Dataset ส่วนข้อมูลที่มาจาก Biased Dataset ให้พิจารณาค่าตัวอื่นๆ แทน หาได้จากสูตร

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+PN+TN} \times 100$$

2) Recall เหมาะสำหรับเรื่องเกี่ยวกับสุขภาพ การบิน การก่อการร้าย ซึ่งไม่ต้องการให้มี False Negative สูง หาได้จากสูตร

$$\text{Recall} = \frac{TP}{TP+FN}$$

3) Precision เหมาะสำหรับ การตัดสินความผิด หรือ Search Engine ที่ต้องการค่า False Positive ต่ำ หาได้จากสูตร

$$\text{Precision} = \frac{TP}{TP+FP}$$

4) F1 เป็นค่าเฉลี่ยของ Recall + Precision

5. การท่ายทอดสิ่งที่ได้จากข้อมูล ได้เรียนรู้อะไรบ้าง ผลลัพธ์ที่ได้นั้นสมเหตุสมผลหรือไม่ และสามารถที่จะเล่าเรื่องได้อย่างมีประสิทธิผลหรือไม่ (ตั้งชูพงศ์, 2565; Khan, S., & Alqahtani, S., 2020)

ส่วน Simplilearn (2018) ได้เพิ่มขั้นตอนการเตรียมข้อมูล (Data Preparation) เพื่อทำความสะอาดข้อมูลและแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมจะนำไปใช้งาน (Data Cleaning and Transformation) และขั้นตอนสุดท้าย การนำไปใช้งานจริง (Deploy and Maintain) ด้วย

2.5 ประโยชน์ของการทำวิเคราะห์ข้อมูล

ประโยชน์ของการทำวิเคราะห์ข้อมูล (Big Data Analysis) ในภาคการศึกษาก่อนให้เกิดประโยชน์อย่างมากต่อตัวนักศึกษาและสถานศึกษา ทำให้มีการปฏิรูปวิธีการบริหารการศึกษา (Joshi, N., 2017; Spear, E., n.d.; Matthews, K., 2018; GovTech., 2015) ได้แก่

1. การวิเคราะห์ข้อมูลขนาดใหญ่จะช่วยทำให้ผลการเรียนของนักศึกษาดีขึ้น ที่ผ่านมาจะมีการประเมินผลการศึกษาจากการทำงานที่ได้รับมอบหมายและการสอบเท่านั้น แต่ชีวิตของนักศึกษาจะมีการข้อมูลที่เกิดขึ้นกับนักศึกษาแต่ละคนจำนวนมาก ไม่ว่าจะเป็น ระยะเวลาในการทำข้อสอบแต่ละข้อ ไม่ทำข้อไหนบ้าง ศึกษาข้อมูลจากแหล่งใดในการเตรียมตัวสอบ ตัวอย่างข้อมูลแบบนี้ถ้าสามารถนำมาใช้ในการแจ้งนักศึกษาได้ในทันทีถ้าใช้ข้อมูลขนาดใหญ่ รวมถึงการแนะนำนักศึกษาคนอื่นๆ ถึงแนวทางการศึกษาแต่ละวิชาให้ประสบความสำเร็จตามทักษะของแต่ละคนอย่างเหมาะสม หรือปรับลดหรือเพิ่มการให้บริการอื่นๆ ที่เกี่ยวข้องที่นักศึกษาเข้าไปใช้บริการ เช่น การใช้บริการห้องสมุด การใช้บริการเครือข่ายอินเตอร์เน็ต ทำให้ประหยัดค่าพลังงานและค่าใช้จ่ายที่เกี่ยวข้อง เป็นต้น

2. การใช้ข้อมูลขนาดใหญ่เพื่อปรับปรุงหลักสูตร ให้เหมาะสมกับผู้เรียนแต่ละคนซึ่งอาจจะมีทักษะในการเรียนรู้ที่ต่างกัน ให้สามารถเรียนได้ตามทักษะที่นักเรียนมี หรือ มีหลักสูตรไหนบ้างที่นักศึกษาลดลง อาจารย์คนไหนที่มีจำนวนนักศึกษาน้อยลง อาจจะต้องมีการลดจำนวนห้องเรียนลงเพื่อประหยัดค่าใช้จ่าย หรือการเปิดให้นักศึกษาประเมินการเรียนการสอนของอาจารย์ประจำวิชาเมื่อสิ้นสุดการสอนแต่ละภาคการศึกษา ภาควิชาก็สามารถเอาข้อมูลเหล่านี้มาปรับปรุงการเรียนการสอบหรือ สิ่งอำนวยความสะดวกต่างๆ ได้อย่างเหมาะสม

3. การใช้ข้อมูลขนาดใหญ่เพื่อลดการตกออกหรือลาออกเนื่องจากเรียนไม่ได้ โดยการใช้การวิเคราะห์ข้อมูลแบบพยากรณ์จะทำให้สามารถป้องกันการตกออกของนักศึกษาที่มีต้องการความช่วยเหลือในช่วงเวลาตั้งแต่เนิ่นๆ ทำให้มีระบบการแจ้งเตือนจากการวิเคราะห์ข้อมูลขนาดใหญ่

4. การใช้ข้อมูลขนาดใหญ่เพื่อใช้วางแผนในการรับนักศึกษาให้ได้ตามเป้าที่วางไว้ เช่นการวิเคราะห์ปัจจัยที่มีผลกระทบต่อขั้นตอนการรับสมัคร ใช้ในการวางกลยุทธ์ และการให้ทุนที่เหมาะสม หรืออาจเพิ่มการเรียนออนไลน์ตามผลการวิเคราะห์ข้อมูลขนาดใหญ่ หรือข้อมูลที่นักศึกษาประสบ

ความสำเร็จในกิจกรรมอื่นๆ เช่น นักกีฬาที่มีชื่อเสียง ก็ช่วยทำให้มหาวิทยาลัยนำมาใช้ประกอบการประชาสัมพันธ์มหาวิทยาลัยได้

การใช้ข้อมูลขนาดใหญ่ในสถาบันการศึกษาจะช่วยให้สามารถแข่งขันกับสถาบันอื่นๆ ได้ด้วยการแก้ปัญหาหรือให้ความสำคัญต่อสิ่งที่นักศึกษาต้องการ รวมถึงการเข้าถึงข้อมูลเชิงลึกที่อาจจะไม่เคยรับทราบมาก่อน ข้อมูลขนาดใหญ่ในวิทยาลัยหรือมหาวิทยาลัยมีคุณค่าสูงต่อการเรียนการสอน สามารถเปลี่ยนโหมดการเรียนการสอบในสถานศึกษาได้ (Meng, L.,2014)

Khan, S., & Alqahtani, S. (2020) และ Alkhalil, A., Abdallah, M. A., Alogali, A., & Aljaloud, A. (2021). ได้สรุปการวิเคราะห์ข้อมูลในสถาบันการศึกษาจะนิยมใช้การวิเคราะห์ดังต่อไปนี้

1. Clustering
2. Relationship Mining
3. Visualization
4. Classification
5. Association Analysis
6. Regression

2.6 งานวิจัยที่เกี่ยวข้อง

Tuaha, S., Siddiqui, I.F., & Ali Arain, Q. (2019) ได้ศึกษาเรื่อง Analyzing Students' Academic Performance through Educational Data Mining. พบว่า ผลงานทางวิชาการคือเสาหลักของความสำเร็จในอนาคตของนักศึกษาและเป็นส่วนสำคัญที่สถาบันการศึกษาทั่วโลกให้ความสนใจ ในปัจจุบันมีการใช้ระบบการจัดการการเรียนการสอบแบบออนไลน์ (e–Learning Management System) เพิ่มมากขึ้นอย่างรวดเร็ว ประเทศที่พัฒนาแล้วได้เปลี่ยนระบบการศึกษาไปเป็นระบบแบบเต็มหรือกึ่งอัตโนมัติเพราะว่าระบบเหล่านี้จะมีการสร้างข้อมูลจำนวนมหาศาลที่ซ่อนองคค์ความรู้หรือรูปแบบที่สามารถใช้สร้างองค์ความรู้ที่มีประโยชน์เพื่อช่วยนักศึกษาปรับปรุงผลการเรียนและความสำเร็จในการศึกษา ในการศึกษานี้ได้แนะนำโมเดลการวัดความรู้ความสามารถทางวิชาการของนักศึกษาด้วยปัจจัยใหม่ๆ ที่เกี่ยวข้องกับการตรงต่อเวลาของการเข้าเรียนและการมีส่วนร่วมของผู้ปกครองในขบวนการเรียนรู้ มีการใช้โมเดลเพื่อวัดประสิทธิภาพทางด้านการศึกษาของนักศึกษาที่เป็นอัลกอริทึมการจำแนกประเภท 3 โมเดลคือ Decision Tree, Naïve Bayes และ Artificial Neural Network (ANN) ผลการศึกษาพบว่าปัจจัยหรือฟีเจอร์การตรงต่อเวลาของการเข้าเรียนและการมีส่วนร่วมของผู้ปกครองในขบวนการเรียนรู้มีผลกระทบสูงต่อความสำเร็จในการเรียน

ของนักศึกษา โมเดลให้ค่าความถูกต้อง (Accuracy) ในระดับดี โดยถ้าเพิ่มฟีเจอร์เหล่านี้จะทำให้ได้ผลลัพธ์ดีขึ้น 10 - 15% เมื่อเทียบกับตอนที่เอาฟีเจอร์เหล่านี้ออกจากโมเดล ทั้งนี้ ANN เป็นอัลกอริทึมการจำแนกประเภทที่มีประสิทธิภาพสูงสุดมีความแม่นยำ 78.1 ตามด้วย Decision Tree 71.1 และ Naïve Bayes 67.6

Viloria, A., Alexa, S. N., Hugo Hernández Palma, Núñez, W. N., & Núñez, L. N. (2020) ได้ศึกษาเรื่อง Using Big Data to Determine Potential Dropouts in Higher Education พบว่าในมหาวิทยาลัยจะมีปัญหาที่คล้ายกันคือการตกออกของนักศึกษาไม่เฉพาะละตินอเมริกาแต่ในประเทศที่พัฒนาแล้วก็เจอปัญหานี้ ปัจจัยที่ใช้วัดคุณภาพของการศึกษาหรือความสำเร็จของสถาบันการศึกษาคือเวลาที่นักศึกษาใช้ในการสำเร็จการศึกษา ซึ่งเกี่ยวข้องโดยตรงการการตกออก ซึ่งภาพรวมของทั้งโลกจะอยู่ที่ร้อยละ 42 ในประเทศสหรัฐอเมริการจะมีการตกออกประมาณร้อยละ 30 ทำให้มีการสูญเสียงบประมาณว่า 9 ล้านล้านเหรียฐสหรัฐสำหรับนักศึกษากลุ่มที่ตกออก การพ้นสภาพนักศึกษาไม่เพียงแต่มีผลกระทบต่อคุณภาพการศึกษาและเศรษฐกิจของประเทศแต่ยังส่งผลต่อการพัฒนาสังคม เพราะสังคมคาดหวังว่าประชากรกลุ่มนี้จะเกื้อหนุนสังคมไม่ว่าจะเป็นการสร้างนวัตกรรม องค์ความรู้ และค้นคว้าวิจัยทางวิทยาศาสตร์ การศึกษาจะใช้สถิติการเรียนรู้ขั้นพื้นฐานในการพยากรณ์ความเป็นไปได้ในการตกออกจากข้อมูลส่วนตัวของนักศึกษาและข้อมูลทางการศึกษา

Tasmin, R., Muhammad, R. N., & Nor Aziati, A.H. (2020) ได้แนะนำการทำการวิเคราะห์ข้อมูลทางการศึกษาดังนี้

- ติดตามผลการเรียนของนักศึกษา การวิเคราะห์ข้อมูลแบบเรียลไทม์จะช่วยทำให้ทราบว่านักศึกษาคนไหนกำลังต้องการความช่วยเหลือ ได้รับความกดดันหรือกำลังอยู่ในความเสี่ยงในการตกออก ทำให้เข้าใจนักศึกษาและช่วยให้นักศึกษามีผลการเรียนที่ดีขึ้น

- เพิ่มอัตราการคงอยู่ของนักศึกษาด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ เช่น การวิเคราะห์การเรียนรู้ที่เหมาะสมสำหรับนักศึกษาแต่ละคนด้วยระบบการจัดการวิชาเรียนที่มีแบบการเรียนที่เหมาะสมที่จะสามารถค้นหาได้ว่านักศึกษาคนไหนกำลังประสบปัญหาในการเรียน ทำให้มีการแจ้งเตือนได้ล่วง หน้าทำให้สามารถให้คำปรึกษากับนักศึกษาได้ทันเวลา

- การให้ทุนการศึกษาแก่นักศึกษาที่มีความจำเป็นจริงๆ ด้วยการวิเคราะห์ข้อมูลส่วนตัวของนักศึกษาว่าคนไหนยังไม่ได้ชำระเงินค่าเรียนบ้าง

- กิจกรรมนอกชั้นเรียนช่วยพัฒนาสุขภาพจิตของนักศึกษา ไม่ว่าจะเป็นการติวนอกชั้นเรียนและการให้คำปรึกษา การออกกำลังกาย หรือกิจกรรมทางสังคมต่างๆ การแนะแนวอาชีพ เป็นต้น

K P Prakash and K. Selvakumari. (2021) ได้ศึกษาเรื่อง Mathematical Modelling and Big-Data Analytics for Student Performance. พบว่า การประเมินผลการเรียนของนักศึกษามีประโยชน์อย่างมากต่ออาจารย์และผู้บริหารในการปรับปรุงการเรียนการสอนและเป้าหมายหรือขบวนการเรียนของนักศึกษา ช่วยให้ระบบการศึกษาสามารถติดตามสอดส่องผลการเรียนของนักศึกษาอย่างเป็นระบบด้วยการใช้เทคโนโลยีในปัจจุบัน ข้อมูลนักศึกษาที่ทำกิจกรรมการเรียนในแต่ละวันหรือภาคการศึกษาจะถูกนำมาประเมินผลการเรียนของนักศึกษา เมื่อเรียนจบนักศึกษาจะได้รับทักษะที่จำเป็นในการสำเร็จการศึกษา การเรียนการสอนที่ปรับให้เหมาะสมกับนักศึกษาแต่ละคน (Adaptive Testing Model) จะถูกใช้ตั้งแต่ปีแรก ถ้ามีการดำเนินการแบบนี้ทุกภาคการศึกษาแนะนำให้ทำอย่างน้อย 3-4 ครั้งต่อปี จะช่วยปรับปรุงประสิทธิภาพการเรียนของนักศึกษาได้รวมถึงอาจารย์ก็สามารถปรับปรุงวิธีการสอน

A Yunita et al (2021). ได้ศึกษาเรื่อง Research Review on Big Data Usage for Learning Analytics and Educational Data Mining: A Way Forward to Develop an Intelligent Automation System. พบว่า การนำเทคโนโลยีดิจิทัลมาใช้งานและการพัฒนาเทคโนโลยีสารสนเทศโดยเฉพาะอย่างยิ่ง Artificial Intelligence (AI) จะถูกนำมาใช้งานในทุกภาคส่วน ทำให้เกิดข้อมูลมหาศาล และก่อให้เกิดโอกาสในการนำข้อมูลเหล่านี้มาวิเคราะห์ซึ่งก็รวมถึงในระบบการศึกษา การศึกษาครั้งนี้จะทบทวนงานวิจัยที่เกี่ยวข้องการข้อมูลขนาดใหญ่ในแวดวงการศึกษาและอธิบายถึงทิศทางในอนาคต ด้วยการใช้ Kitchenham's technique เพื่อเลือกและจัดกลุ่มวรรณกรรมออกเป็นประเภทของข้อมูล วิธีการ ประเภทของการวิเคราะห์ และการใช้งานในการวิเคราะห์ข้อมูลการเรียนรู้ ผลการวิจัยพบว่า การวิเคราะห์ข้อมูลขนาดใหญ่สำหรับการศึกษาส่วนใหญ่จะมีวัตถุประสงค์เพื่อปรับปรุงขบวนการเรียนรู้ วิเคราะห์พฤติกรรมของผู้เรียน เพิ่มอัตราการคงอยู่ของนักศึกษา และประเมินผลของนักศึกษาที่ใช้ระบบ MOOCs และระบบการจัดการเรียนการสอน ทิศทางในอนาคตที่สำคัญได้แก่ 1) การสร้างชุดข้อมูลขนาดใหญ่ที่เปิดให้ทุกคนเข้าถึงได้ รวมถึงการเตรียมข้อมูลและปัญหาของชุดข้อมูลที่ไม่สมดุลย์กัน (Imbalanced Dataset) 2) วิเคราะห์ข้อมูลจากกิจกรรมที่เก็บไว้ในประวัติการเรียน เพื่อหาองค์ความรู้และข้อมูลเชิงลึกจากพฤติการการเรียนออนไลน์ไม่เพียงเฉพาะในมุมมองของผู้เรียนแต่รวมไปถึงกิจกรรมต่างๆ ของผู้สอนด้วย 3) ออกแบบระบบการวิเคราะห์ข้อมูลอัตโนมัติเพื่อทำการอธิบายข้อมูล ทำนาย และให้คำแนะนำในการเรียนรู้ด้วย สรุป การเปิดรับข้อมูลขนาดใหญ่ในระบบการศึกษาเพื่อวิเคราะห์การเรียนรู้และการทำเหมืองข้อมูลทางการศึกษาเป็นพื้นที่สำหรับงานวิจัยที่มีประโยชน์ต่อการศึกษาอย่างมาก

บทที่ 3

วิธีดำเนินการศึกษาวิจัย

ในการศึกษาปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยขอนแก่นทั้งนี้ผู้วิจัยได้ดำเนินการวิจัยตามลำดับ ดังนี้

3.1 ประชากรและกลุ่มตัวอย่าง

3.2 เครื่องมือที่ใช้ในงานวิจัย

3.1 ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการศึกษาครั้งนี้ เป็นการวิเคราะห์ข้อมูลในฐานข้อมูลระบบทะเบียนนักศึกษาระดับปริญญาตรีที่เข้าศึกษาด้วยวิธีการคัดเลือกแบบโควตาที่เข้าศึกษาในปีการศึกษา 2561 จำนวน 1,792 คน ซึ่งประกอบไปด้วยข้อมูลพื้นฐานของนักศึกษา ได้แก่ เพศ รายได้ของครอบครัว การประกอบอาชีพของผู้ปกครอง ขนาดของโรงเรียน เกรด ม. ปลาย เกรดในการเรียนในชั้นปีที่ 1 คะแนนวิชาหลักที่สาขากำหนดในการคัดเลือกแข่งขัน ลำดับของสาขาที่สอบได้ ข้อมูลการเข้าใช้บริการห้องสมุด ข้อมูลการเข้าใช้อินเตอร์เน็ต และข้อมูลการประเมินการเรียนการสอน โดยปฏิบัติตามพระราชบัญญัติข้อมูลส่วนบุคคลอย่างเคร่งครัด

3.2 เครื่องมือที่ใช้ในการวิจัย

การศึกษาครั้งนี้เป็นการวิเคราะห์สถิติเชิงอนุมาน (Inferential Statistics Analysis) โดยใช้โปรแกรม Python ด้วยวิธี Logistic Regression และ Decision Tree ผ่าน Google Collaboratory ซึ่งมีขั้นตอนการวิเคราะห์ดังนี้

3.2.1 เตรียมข้อมูล

3.2.2 Import Library ที่จะใช้งานและโหลดข้อมูล

3.2.3 สำรวจข้อมูล

3.2.4 ฝึกโมเดล

3.2.5 ทดสอบความแม่นยำของโมเดล

3.2.6 นำไปใช้พยากรณ์

บทที่ 4

ผลการศึกษา

การวิเคราะห์ข้อมูลปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยขอนแก่น ผู้วิจัยได้นำข้อมูลจากฐานข้อมูลรับเข้าศึกษาและฐานข้อมูลทะเบียนนักศึกษาที่เข้าศึกษาด้วยวิธีรับตรงประจำปีการศึกษา 2561 จำนวน 1,641 รายการ เมื่อทำความสะอาดข้อมูลโดยลบข้อมูลที่ไม่สมบูรณ์ หรือ ข้อมูลที่มีค่าผิดปกติแล้วทำให้มีข้อมูลสำหรับวิเคราะห์ทั้งสิ้น 1,512 รายการ วิเคราะห์ด้วยวิธีการทางสถิติเชิงอนุมาน (Inferential Statistics Analysis) โดยใช้โปรแกรม Python ด้วยวิธี Logistic Regression และ Decision Tree ผ่าน Google Collaboratory ด้วยการนำเสนอผลการวิเคราะห์ข้อมูลดังต่อไปนี้

4.1 การเตรียมข้อมูล

ข้อมูลที่นำมาวิเคราะห์ครั้งนี้ประกอบไปด้วยปัจจัย (Features) ดังต่อไปนี้

1) GRADUATION
   การสำเร็จการศึกษา - 0 = ไม่สำเร็จการศึกษา 1 = สำเร็จการศึกษา
2) GPA1
   GPA ของนักศึกษาในภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1
3) GPA2
   GPA ของนักศึกษาในภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1
4) SELECTIONORDER
   ลำดับในการเลือกสาขาที่ผ่านการคัดเลือก อันดับที่ 1 - 4
5) SEX
   เพศ M = ชาย F = หญิง
6) MYRANK
   คือลำดับที่สอบผ่านการคัดเลือกโดยเรียงจากลำดับที่ 1 คือลำดับที่ดีที่สุด
7) ENG
   คือคะแนนวิชาภาษาอังกฤษ
8) PARENTEDU
   คือระดับการศึกษาของผู้ปกครอง

9) PARENTCAREER

คือ อาชีพของผู้ปกครอง

10) PARENTINCOME

คือ รายได้ของผู้ปกครองต่อเดือน

11) TEACHEREVAL

ค่าเฉลี่ยของผลการประเมินอาจารย์ผู้สอนทุกวิชา มีค่าตั้งแต่ 1 - 5

12) CLASSEVAL

ค่าเฉลี่ยของผลการประเมินวิชาที่เรียนทุกวิชา มีค่าตั้งแต่ 1 - 5

4.2 Import Library ที่จะใช้งานและโหลดข้อมูล

4.2.1 ทำการโหลด Library ที่ใช้ในการวิเคราะห์ข้อมูลใน Google Colab

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns; sns.set()
```

4.2.2 โหลดข้อมูลที่ต้องการวิเคราะห์ที่เตรียมไว้ในไฟล์ MS Excel

```
# Supress warnings
import warnings
warnings.filterwarnings('ignore')

# Importing the Data Set
import pandas as pd
df = pd.read_excel('Graduation2561_2.xlsx')
```

4.2.3 ทำการวิวข้อมูลตัวอย่าง 5 รายการแรกด้วยคำสั่ง head(5)

```
df.head(5)
```

	DEFAULTSEX	PROGRAMID	FACULTYID	GRADUATED	MYRANK	SCORE	SELECTIONORDER	THAI	SOC	ENG	...
0	F	312345140160	45	1	13	26.950	4	36.0	28.0	18.75	...
1	F	313201001160	32	1	55	38.075	2	62.0	40.0	31.25	...
2	F	312109501160	21	1	21	44.988	1	60.0	44.0	38.75	...
3	M	311101701160	11	1	32	39.388	3	70.0	46.0	52.50	...
4	F	310207101160	2	1	10	39.800	2	70.0	42.0	52.50	...

5 rows × 23 columns

รูปภาพที่ 2 แสดงข้อมูลข้อมูล 5 รายการแรกด้วยคำสั่ง head(5)

4.2.4 ใช้คำสั่ง info() สำหรับสำรวจข้อมูลของ Dataset

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1641 entries, 0 to 1640
Data columns (total 23 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   DEFAULTSEX      1641 non-null   object
 1   PROGRAMID       1641 non-null   int64
 2   FACULTYID       1641 non-null   int64
 3   GRADUATED       1641 non-null   int64
 4   MYRANK          1641 non-null   int64
 5   SCORE           1641 non-null   float64
 6   SELECTIONORDER  1641 non-null   int64
 7   THAI            1629 non-null   float64
 8   SOC             1628 non-null   float64
 9   ENG             1641 non-null   float64
 10  MATH1           1112 non-null   float64
 11  PHYSICS         739 non-null    float64
 12  CHEM            787 non-null    float64
 13  BIO             796 non-null    float64
 14  MATH2           479 non-null    float64
 15  SCI             768 non-null    float64
 16  GPA1            1641 non-null   float64
 17  GPA2            1633 non-null   float64
 18  PARENTEDU       1641 non-null   int64
 19  PARENTCAREER    1641 non-null   int64
 20  PARENTINCOME    1568 non-null   float64
 21  TEACHEREVAL     1641 non-null   float64
 22  CLASSEVAL       1641 non-null   float64
dtypes: float64(15), int64(7), object(1)
memory usage: 295.0+ KB
```

รูปภาพที่ 3 ผลของการใช้คำสั่ง info() สำหรับสำรวจข้อมูลของ Dataset

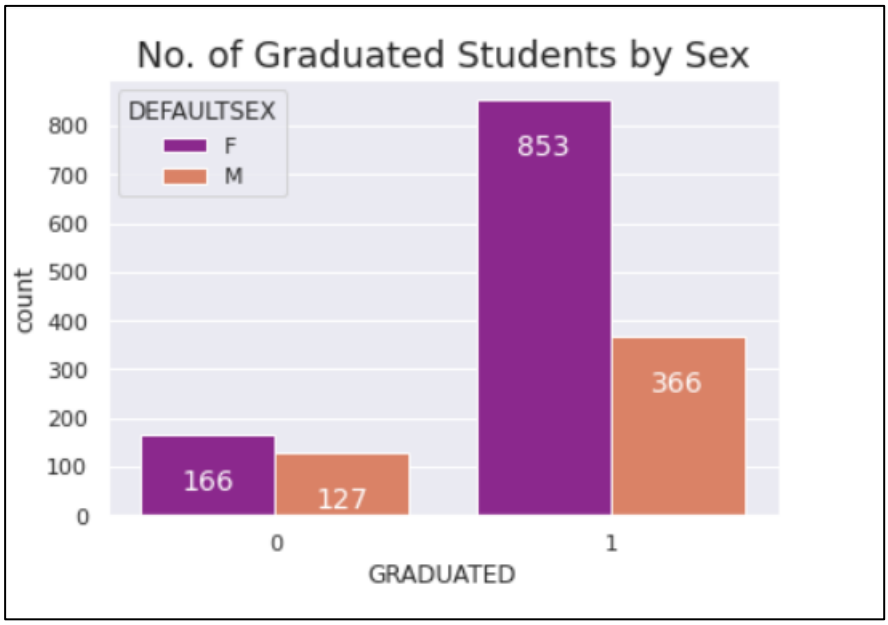
4.2.5 จากนั้นความสะอาดข้อมูลรายได้ของผู้ปกครองที่มีค่าสูงจากกลุ่มจำนวนมากออก

```
def remove_outlier(df_in, col_name):
    q1 = df_in[col_name].quantile(0.25)
    q3 = df_in[col_name].quantile(0.75)
    iqr = q3-q1 #Interquartile range
    fence_low  = q1-1.5*iqr
    fence_high = q3+1.5*iqr
    df_out = df_in.loc[(df_in[col_name] > fence_low) &
                       (df_in[col_name] < fence_high)]
    return df_out

df = remove_outlier(df, 'PARENTINCOME')
```

4.3 สำรวจข้อมูล

4.3.1 เริ่มจากการแสดงข้อมูลจำนวนผู้สำเร็จการศึกษาแยกตามเพศ



รูปภาพที่ 4 แสดงข้อมูลจำนวนผู้สำเร็จการศึกษาแยกตามเพศ

จากรูปภาพที่ 4 จะเห็นว่าจำนวนผู้สำเร็จการศึกษาคิดเป็นร้อยละ 80.62 โดยผู้ที่ไม่สำเร็จการศึกษาที่เป็นเพศชายคิดเป็นร้อยละ 25.76 ในขณะที่ผู้หญิงที่ไม่สำเร็จการศึกษาคิดเป็นร้อยละ 16.29

4.3.2 จากนั้นทำการสำรวจข้อมูลจำนวนผู้สำเร็จการศึกษาแยกตามเพศพร้อมทั้งแสดงผลการเรียน (GPA) ของภาคศึกษาที่ 1 และ 2

```
X=df.groupby('DEFAULTSEX')['GRADUATED'].count()
X
```

จะพบว่าเพศชายสำเร็จการศึกษาทั้งสิ้น 493 คน เพศหญิง 1,019 คน แล้วทำการแยกจำนวนผู้สำเร็จการศึกษาตามเพศและคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 และ 2 ของการเรียนในชั้นปีที่ 1

```
X=df.groupby(['GRADUATED','DEFAULTSEX'])['GPA1','GPA2'].mean()
X
```

		GPA1	GPA2
GRADUATED	DEFAULTSEX		
0	F	2.355181	1.726375
	M	2.007323	1.336508
1	F	2.848781	2.834138
	M	2.684754	2.634754

รูปภาพที่ 5 แสดง GPA เฉลี่ยของนักศึกษาแยกตามการสำเร็จการศึกษาและเพศ

4.3.3 จากนั้นทำการตรวจสอบคุณภาพของข้อมูลที่นำเข้ามวิเคราะห์ด้วยคำสั่ง isna().sum()

```
df.isna().sum()
```

```
DEFAULTSEX          0
PROGRAMID           0
FACULTYID           0
GRADUATED           0
MYRANK              0
SCORE               0
SELECTIONORDER      0
THAI               10
SOC                11
ENG                 0
MATH1             483
PHYSICS           827
CHEM              783
BIO               776
MATH2            1076
SCI               810
GPA1                0
GPA2                7
PARENTEDU           0
PARENTCAREER        0
PARENTINCOME        0
TEACHEREVAL         0
CLASSEVAL           0
dtype: int64
```

รูปภาพที่ 6 แสดงข้อมูลแต่ละฟิลด์ที่มีค่าว่าง

จะเห็นว่าคะแนนวิชาที่ใช้ในการสอบคัดเลือกจะมีเฉพาะคะแนนวิชาภาษาอังกฤษที่ทุกสาขาต้องใช้ ส่วนวิชาอื่นๆ ไม่ได้ใช้ในบางสาขา ส่วนคะแนนผลการเรียนภาคการศึกษาที่ 2 มีค่าว่างจำนวน 7 รายการ จึงต้องทำการตัดปัจจัยที่เป็นคะแนนวิชาสอบออกยกเว้นคะแนนวิชาภาษาอังกฤษ และตัดข้อมูลรายการที่มีค่าว่างในปัจจัยผลการเรียนภาคเรียนที่ 2 ออก

4.4 การแก้ปัญหา Multicollinearity

ในการวิเคราะห์ข้อมูลต้องทำการแปลงข้อมูลในปัจจัยเพศให้เป็นข้อมูลประเภทกลุ่ม โดย 0 คือ เพศหญิง และ 1 คือ เพศชาย เพื่อลดปัญหาเพศหญิงและชายที่มีความสัมพันธ์กันสูง

```
# to avoid multicollinearity
df2 =pd.get_dummies(df,columns=['DEFAULTSEX'],
                    prefix='',prefix_sep='',drop_first = True)
```

4.4.1 ทำการตัดปัจจัยที่ไม่นำมาวิเคราะห์ออกซึ่งก็จะมีปัจจัย PROGRAMID, FACULTYID, THAI, SOC,MATH1, PHYSICS, CHEM, BIO, MATH2, SCI ด้วยคำสั่ง drop()

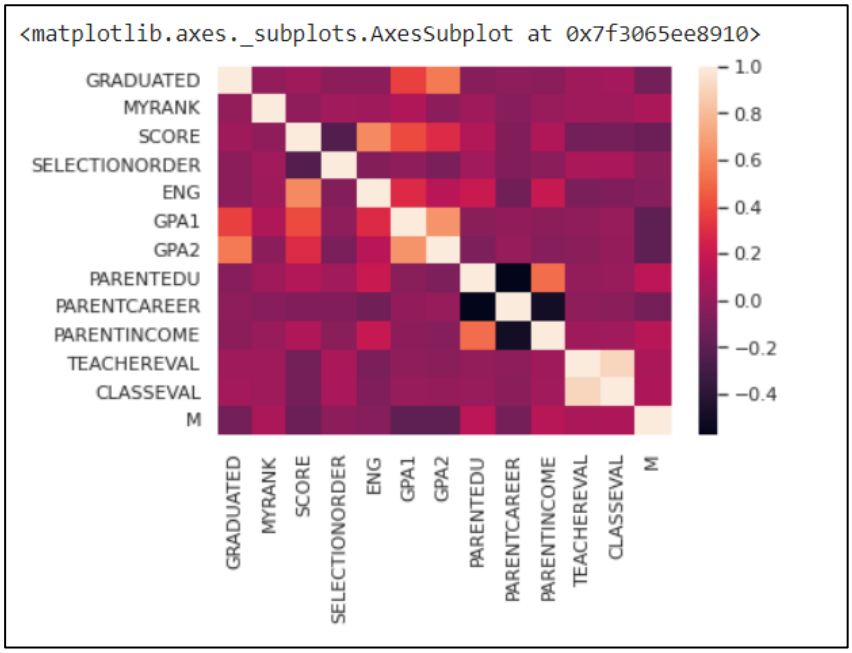
```
df2.drop(['PROGRAMID','FACULTYID','THAI', 'SOC','MATH1', 'PHYSICS',
        'CHEM','BIO', 'MATH2', 'SCI'], axis = 1, inplace = True)
```

	GRADUATED	MYRANK	SCORE	SELECTIONORDER	ENG	GPA1	GPA2	PARENTEDU	PARENTCAREER	PARENTINCOME	TEACHEREVAL	CLASSEVAL	M
0	1	13	26.950	4	18.75	1.92	2.29	7	3	38000.0	4.39	4.23	0
1	1	55	38.075	2	31.25	2.50	2.78	7	1	50000.0	3.57	3.62	0
2	1	21	44.988	1	38.75	3.08	2.87	1	7	9500.0	4.10	4.23	0
3	1	32	39.388	3	52.50	2.78	3.16	4	6	15000.0	3.76	3.67	1
4	1	10	39.800	2	52.50	2.58	2.41	6	1	40000.0	4.81	4.62	0
5	1	1	60.350	2	47.50	3.45	2.98	3	5	15000.0	3.96	3.55	0
6	1	3	57.400	1	70.00	3.61	3.50	5	5	20000.0	4.85	4.74	1
7	1	30	54.488	2	46.25	3.00	2.92	5	5	10000.0	3.58	3.35	0
8	1	2	62.250	1	67.50	2.93	3.71	3	7	8268.0	3.26	3.29	0
9	1	33	28.300	3	36.25	1.36	2.28	6	1	40000.0	3.59	3.57	0

รูปภาพที่ 7 แสดงข้อมูลที่ตัดเอาปัจจัยที่ไม่นำมาวิเคราะห์ออกแล้ว

4.4.2 จากนั้นสร้าง Heatmap แสดงความสัมพันธ์ (Correlation) ของข้อมูลในภาพรวม ด้วยคำสั่ง headmap() ของ sns

```
sns.heatmap(df2.corr())
```



รูปภาพที่ 8 แสดงความสัมพันธ์ (Correlation) ของข้อมูลในภาพรวม

4.4.3 แล้วทำการตัดข้อมูลที่มีค่าคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 จำนวน 7 รายการที่มีค่าว่างออก ด้วยคำสั่ง dropna()

```
df2.dropna(inplace = True)
```

```
GRADUATED          0
MYRANK             0
SCORE              0
SELECTIONORDER     0
ENG                0
GPA1               0
GPA2               7
PARENTEDU          0
PARENTCAREER       0
PARENTINCOME       0
TEACHEREVAL        0
CLASSEVAL          0
M                  0
dtype: int64
```

รูปภาพที่ 9 แสดง GPA2 ที่มีค่าว่าง

4.5 การฝึกโมเดล

4.5.1 เริ่มต้นจากการแยกตัวแปรต้นและตัวแปรตามออกจากกันก่อน ซึ่งตัวแปรตามคือ การสำเร็จการศึกษา (GRADUATED)

```
x_data = df2.drop(columns=['GRADUATED'])
y_data = df2['GRADUATED']
```

4.5.2 ทำการฝึกโมเดลด้วยการแยกข้อมูลออกเป็น 2 ชุด สำหรับฝึกโมเดล (Train Set, 70%) และ สำหรับทดสอบโมเดล (Test Set, 30%)

```
from sklearn.model_selection import train_test_split
x_training_data, x_test_data, y_training_data,
        y_test_data = train_test_split(x_data, y_data, test_size = 0.3,
        random_state=0)
```

แล้วทำการลดสเกลปัจจัยที่ใช้เพื่อให้อยู่ในสเกลเดียวกัน

```
from sklearn.preprocessing import StandardScaler
standard_scaler = StandardScaler()
x_training_data.loc[:,['MYRANK', 'SCORE','ENG','PARENTINCOME',
      'PARENTEDU','PARENTCAREER','TEACHEREVAL','CLASSEVAL'
      ]]=standard_scaler.fit_transform(x_training_data.loc[:,
        ['MYRANK', 'SCORE','ENG','PARENTINCOME','PARENTEDU',
      'PARENTCAREER','TEACHEREVAL','CLASSEVAL']])
```

4.5.2 จากนั้นทำการฝึกโมเดล Linear Regression

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression(random_state=0)
model
```

แล้วทำการฟิตโมเดล

```
model.fit(x_training_data,y_training_data)
```

4.5.3 เมื่อเทรนเสร็จ ก็ทำการทำนายข้อมูลที่เตรียมไว้สำหรับทดสอบโมเดล

```
x_test_data.loc[:,['MYRANK', 'SCORE','ENG','PARENTINCOME','PARENTEDU'
,'PARENTCAREER','TEACHEREVAL','CLASSEVAL']]=standard_scaler.transform
(x_test_data.loc[:,['MYRANK', 'SCORE','ENG','PARENTINCOME','PARENTEDU
','PARENTCAREER','TEACHEREVAL','CLASSEVAL']])
```

เสร็จแล้วก็แสดงผลการการทำนาย

```
predictions = model.predict(x_test_data)
```

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1,
       1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1,
       1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
```

รูปภาพที่ 10 แสดงผลการทำนายการสำเร็จการศึกษา

4.6 ทดสอบความแม่นยำของโมเดล

4.6.1 การประเมินโมเดลในกลุ่ม Classification ซึ่งเป็นโมเดลสำหรับทำนายว่าสำเร็จการศึกษา (1) หรือไม่สำเร็จการศึกษา (0) จะดูผลคะแนนของสิ่งต่อไปนี้

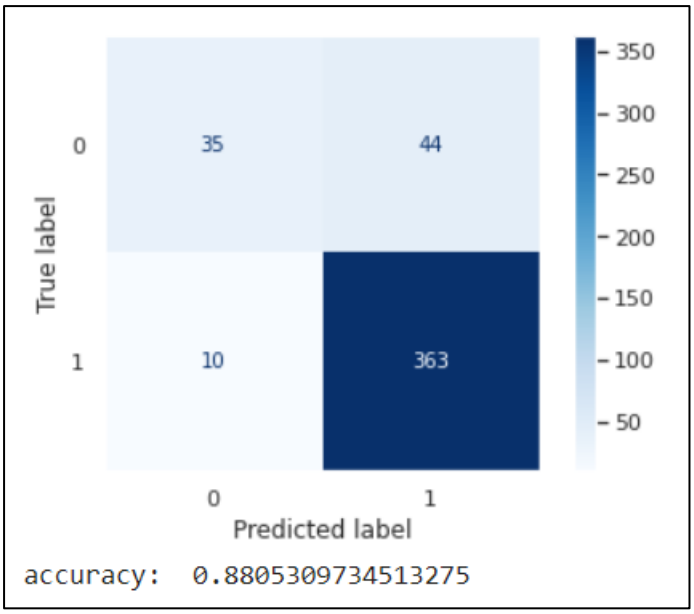
- Confusion Metrix
- Accuracy
- Precision, Recall, F1-Score

ที่มีอยู่ใน sklearn

```
from sklearn.metrics import plot_confusion_matrix, accuracy_score
cm = plot_confusion_matrix(model,x_test_data, y_test_data,cmap="Blues
",values_format='.3g');
plt.grid(None)
plt.show()

y_pred = model.predict(x_test_data)
print("accuracy: ",accuracy_score(y_test_data, y_pred))
```

ซึ่งจะได้ความถูกต้องของการทำนาย 88.05%



รูปภาพที่ 11 แสดง Confussion Metrics ของ Logistict Regression Model ที่ทำนายครั้งแรก

4.6.2 แสดงรายละเอียดประสิทธิภาพของโมเดล

```
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score

print("Accuracy:\t %.3f" %accuracy_score(y_test_data, predictions))
print(classification_report(y_test_data, predictions))
```

จะได้ความแม่นยำในการทำนายของคนที่สำเร็จการศึกษาจริง และผลการทำนาย ได้ 89% และทำนายคนที่ไม่สำเร็จการศึกษาถูกต้อง 78%

```
Accuracy:       0.881
              precision    recall  f1-score   support

           0       0.78      0.44      0.56        79
           1       0.89      0.97      0.93       373

    accuracy                           0.88       452
   macro avg       0.83      0.71      0.75       452
weighted avg       0.87      0.88      0.87       452
```

รูปภาพที่ 12 แสดง Accuracy, Precision, Recall และ F1-score ในการทำนายครั้งแรก

4.6.3 เพื่อให้ผลการทำนายของโมเดลมีความถูกต้องมากยิ่งขึ้น จึงต้องทำการปรับโมเดลเพื่อเพิ่มประสิทธิภาพของโมเดล ด้วยการปรับพารามิเตอร์ต่างๆ (Hyperparameter) โดยเริ่มจากการปรับค่า `Inverse of Regularization Strength (C)` ซึ่งค่าตั้งต้นจะมีค่าเท่ากับ 1.0 การใส่ค่า C น้อยๆ จะหมายถึงเพิ่มความแรง

```
model = LogisticRegression(C=0.1, class_weight=None, dual=False,
      fit_intercept=None,intercept_scaling=1, max_iter=1000,
      multi_class='ovr', n_jobs=1,penalty='l2',random_state=0,
      solver='liblinear', tol=0.0001, verbose=0, warm_start=False)
model.fit(x_training_data, y_training_data)
predictions = model.predict(x_test_data)
print("Accuracy:\t %.3f" %accuracy_score(y_test_data, predictions))
print(classification_report(y_test_data, predictions))
```

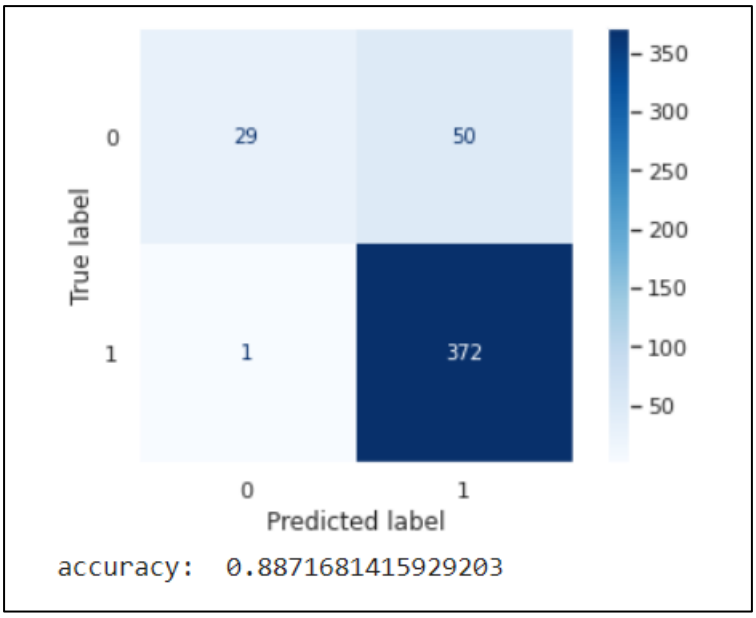
จะเห็นว่าค่า Accuracy เพิ่มขึ้นมาเป็น 88.7% และ ได้ความแม่นยำในการทำนายของคนที่สำเร็จการศึกษาจริง และผลการทำนาย ได้ 88% และทำนายคนที่ไม่สำเร็จการศึกษาถูกต้อง 97% ซึ่งดีขึ้นกว่าเดิม

```
Accuracy:         0.887
              precision    recall  f1-score   support

           0       0.97      0.37      0.53        79
           1       0.88      1.00      0.94       373

    accuracy                           0.89       452
   macro avg       0.92      0.68      0.73       452
weighted avg       0.90      0.89      0.87       452
```

รูปภาพที่ 13 แสดงค่า Accuracy, Precision, Recall, F1-Score หลังปรับ Hyperparameter



รูปภาพที่ 14 แสดงค่า Confussion Metrics หลังปรับ Hyperparameter

4.6.4 จากนั้นลองปรับพารามิเตอร์ด้วยวิธี K-Folds Cross Validation ซึ่งจะสามารถกำหนดให้ทำการทดสอบโมเดลหลายๆ รอบ แล้วนำเอาค่าพารามิเตอร์ที่ดีที่สุดมาใช้ จากตัวอย่างโค้ดข้างล่างจะมีการปรับพารามิเตอร์จำนวน 72 โมเดล (6 x 2 x 3 x 2 = 72 models)

```
parameters = {
    'C': [0.01, 0.1, 0.5, 1, 5, 10],
    'fit_intercept': [True, False],
    'random_state': [0,1,2],
    'solver': ['liblinear', 'lbfgs']
}
from sklearn.model_selection import GridSearchCV
clf = LogisticRegression()
model = GridSearchCV(clf, parameters, cv=5, scoring='accuracy',
      n_jobs=-1)
model.fit(x_training_data, y_training_data)

predictions = model.best_estimator_.predict(x_test_data)
print("Accuracy:\t %.3f" %accuracy_score(y_test_data, predictions))
print(classification_report(y_test_data, predictions))
```

ได้ผลดังนี้

Accuracy:	0.874			
	precision	recall	f1-score	support
0	0.79	0.38	0.51	79
1	0.88	0.98	0.93	373
accuracy			0.87	452
macro avg	0.84	0.68	0.72	452
weighted avg	0.87	0.87	0.86	452

รูปภาพที่ 15 แสดงค่า Accuracy, Precision, Recall, F1-Score หลังปรับค่า K-Folds Cross Validation

4.7 นำไปใช้พยากรณ์

4.7.1 การพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาที่รับด้วยวิธีรับตรง ประจำปีการศึกษา 2561

```
# model coefficient
print(dict(zip(list(x_training_data.columns), model.best_estimator_.c
oef_[0])))
feature_importance_m = abs(model.best_estimator_.coef_[0])
feature_importance_m = 100.0 * (feature_importance_m / feature_import
ance_m.max())
lr_feature_m = pd.DataFrame({'feature':x_training_data.columns,
                             'Score':feature_importance_m})
lr_feature_m.sort_values(by = 'Score', ascending=False).head()
```

	feature	Score
5	GPA2	100.000000
4	GPA1	31.916525
1	SCORE	18.209762
11	M	15.864087
3	ENG	15.347712

รูปภาพที่ 16 แสดงปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษากลุ่มตัวอย่าง

จากตารางที่ 16 พบว่าปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) สูงที่สุด คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ

คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE), เพศ และ ผลคะแนนภาษาอังกฤษ จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา

4.8 ทำนายผลด้วยการใช้ Decision Tree Model

4.8.1 เริ่มจากการนำเข้า DecisionTreeClassifier

```
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier(random_state=0, criterion='entropy')
model
model.fit(x_training_data, y_training_data)
predictions = model.predict(x_test_data)
```

แล้วทำการทำนาย

```
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
print("Accuracy:\t %.3f" %accuracy_score(y_test_data, predictions))
print(classification_report(y_test_data, predictions))
```

ได้ผลดังนี้

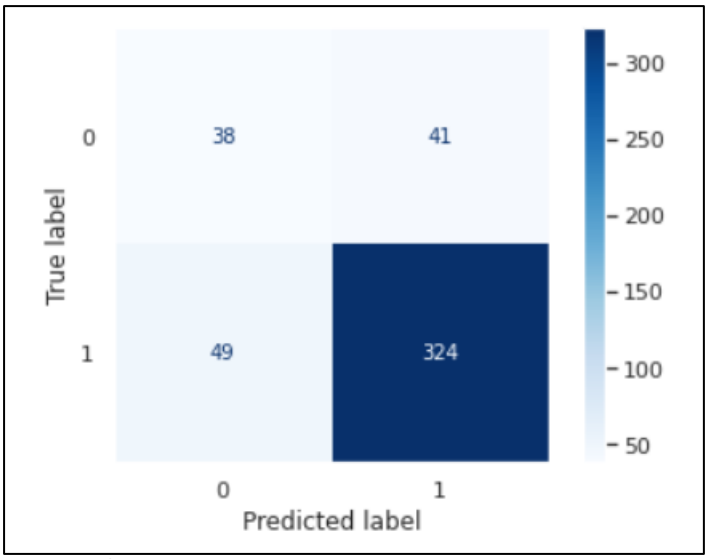
```
Accuracy:        0.801
              precision    recall  f1-score   support

           0       0.44      0.48      0.46        79
           1       0.89      0.87      0.88       373

    accuracy                           0.80       452
   macro avg       0.66      0.67      0.67       452
weighted avg       0.81      0.80      0.80       452
```

รูปภาพที่ 17 แสดงค่า Accuracy, Precision, Recall, F1-Score ของ Decision Tree Model

จากรูปภาพที่ 17 จะได้ค่า Accuracy 80.1% และ ได้ความแม่นยำในการทำนายของคนที่สำเร็จการศึกษาจริง และผลการทำนาย ได้ 89% และทำนายคนที่ไม่สำเร็จการศึกษาถูกต้อง 44%



รูปภาพที่ 18 แสดง Confussion Metrics ของ Decition Tree Model

4.8.2 จากนั้นทำการปรับพารามิเตอร์เพื่อเพิ่มประสิทธิภาพของโมเดล

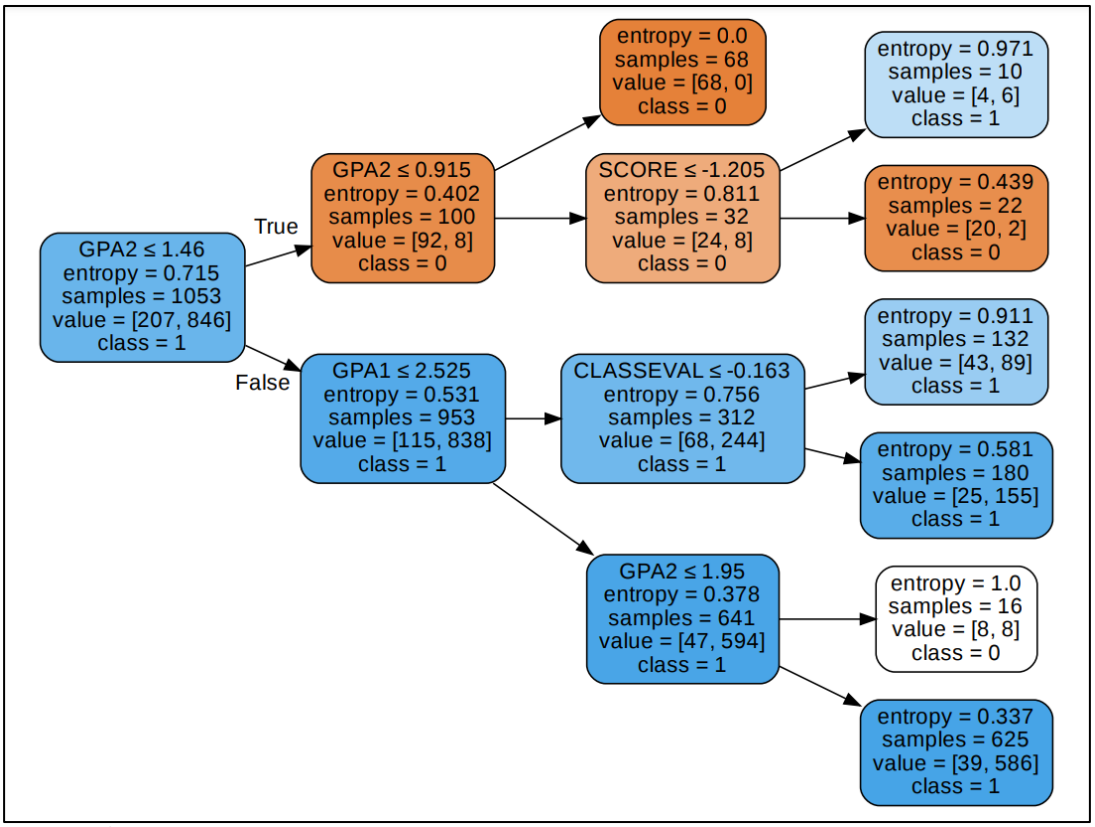
```
model = DecisionTreeClassifier(criterion='entropy',
                               splitter='best',
                               max_depth=3,
                               min_samples_split=2,
                               min_samples_leaf=1,
                               min_weight_fraction_leaf=0.0,
                               max_features=None,
                               random_state=None,
                               max_leaf_nodes=None,
                               min_impurity_decrease=0.0,
                               class_weight=None,
                               ccp_alpha=0.0)
model.fit(x_training_data, y_training_data)
predictions = model.predict(x_test_data)
print("Accuracy:\t %.3f" %accuracy_score(y_test_data, predictions))
print(classification_report(y_test_data, predictions))
```

ทำให้ได้ผลการทำนายที่ดีขึ้นโดยได้ค่า Accuracy 86.5% และ ได้ความแม่นยำในการทำนายของคนที่สำเร็จการศึกษาจริง และผลการทำนาย ได้ 88% และทำนายคนที่ไม่สำเร็จการศึกษาถูกต้อง 71%

Accuracy:	0.865			
	precision	recall	f1-score	support
0	0.71	0.38	0.50	79
1	0.88	0.97	0.92	373
accuracy			0.87	452
macro avg	0.80	0.67	0.71	452
weighted avg	0.85	0.87	0.85	452

รูปภาพที่ 19 แสดงผลของการปรับพารามิเตอร์เพื่อเพิ่มประสิทธิภาพของ Decision Tree Model

เมื่อทำการแสดงผล Decision Tree จะได้ดังภาพข้างล่างนี้



รูปภาพที่ 20 แสดงภาพการทำนายของ Decision Tree Model

จากรูปภาพที่ 20 ของ Decision Tree จะเห็นว่านักศึกษาที่มีผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 มากกว่า 1.46 ขึ้นไปจะสำเร็จการศึกษาทุกคน ส่วนนักศึกษาที่ได้คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 น้อยกว่า 1.46 จะมีโอกาสที่จะไม่สำเร็จการศึกษาสูง

4.8.3 เมื่อทำการดูปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษา จะใช้คำสั่งด้านล่างนี้

```
importances = model.feature_importances_
print(importances)
```

```
'MYRANK': 0.,
'SCORE': 0.02332834,
'SELECTIONORDER': 0.,
'ENG': 0.,
'GPA1': 0.09884622,
'GPA2': 0.83823923,
'PARENTEDU': 0.,
'PARENTCAREER': 0.,
'PARENTINCOME': 0.,
'TEACHEREVAL': 0.,
'CLASSEVAL': 0.0395862,
'M': 0.
```

การพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาที่รับด้วยวิธีรับตรงประจำปีการศึกษา 2561 ด้วย Decision Tree Model พบว่าปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) สูงที่สุด คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE) และ ระดับคะแนนเฉลี่ยการประเมินรายวิชาที่เรียน จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การศึกษาครั้งนี้ต้องการศึกษาว่ามีปัจจัยอะไรบ้างที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรีของมหาวิทยาลัยขอนแก่นบ้าง ได้แก่ 1) ปัจจัยทางด้านส่วนตัว เช่น เพศ รายได้ของครอบครัว การประกอบอาชีพของผู้ปกครอง 2) ปัจจัยทางด้านการศึกษา เช่น คะแนนรวมเกรดเฉลี่ยทั้ง 2 ภาคในการเรียนในชั้นปีที่ 1 คะแนนที่ผ่านการคัดเลือก และ 3) ปัจจัยในการให้บริการของมหาวิทยาลัยขอนแก่น คือ ค่าเฉลี่ยคะแนนของการประเมินการเรียนการสอบของอาจารย์ประจำวิชาและผลการประเมินรายวิชาที่ลงทะเบียนเรียน

ผู้วิจัยได้นำข้อมูลจากฐานข้อมูลรับเข้าศึกษาและฐานข้อมูลทะเบียนนักศึกษา ที่เข้าศึกษาด้วยวิธีรับตรงประจำปีการศึกษา 2561 จำนวน 1,641 รายการ เมื่อทำความสะอาดข้อมูลโดยลบข้อมูลที่ไม่สมบูรณ์ หรือ ข้อมูลที่มีค่าผิดปกติแล้วทำให้มีข้อมูลสำหรับวิเคราะห์ทั้งสิ้น 1,512 รายการ วิเคราะห์ด้วยวิธีการทางสถิติเชิงอนุมาน (Inferential Statistics Analysis) โดยใช้โปรแกรม Python ด้วยวิธี Logistic Regression และ Decision Tree ผ่าน Google Collaboratory

5.1 สรุปผลการวิจัย

ลักษณะทั่วไปของประชากร

5.1.1 จำนวนเพศชายสำเร็จการศึกษาทั้งสิ้น 493 คน เพศหญิง 1,019 คนคิดเป็นร้อยละ 80.62 โดยผู้ที่ไม่สำเร็จการศึกษาที่เป็นเพศชายคิดเป็นร้อยละ 25.76 ในขณะที่ผู้หญิงที่ไม่สำเร็จการศึกษาคิดเป็นร้อยละ 16.29

5.1.2 คะแนนสอบแข่งขันตามเกณฑ์ของสาขา (SCORE) มีค่าเฉลี่ยอยู่ที่ 40.43 มีค่าเบี่ยงเบนมาตรฐาน 9.8407

5.1.3 คะแนนวิชาภาษาอังกฤษที่ได้จากการสอบวัดความรู้วิชาสามัญที่จัดสอบโดย สทศ. (ENG) ค่าเฉลี่ยอยู่ที่ 39.40 มีค่าเบี่ยงเบนมาตรฐาน 12.9010

5.1.4 คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1) มีค่าเฉลี่ยอยู่ที่ 2.68 มีค่าเบี่ยงเบนมาตรฐาน 0.6341 คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) ค่าเฉลี่ยอยู่ที่ 2.54 มีค่าเบี่ยงเบนมาตรฐาน 0.8575

ทั้งนี้นักศึกษาที่สำเร็จการศึกษาที่เป็นเพศชาย จะมีคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1) มีค่าเฉลี่ยอยู่ที่ 2.68 และคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) ค่าเฉลี่ยอยู่ที่ 2.63

นักศึกษาที่สำเร็จการศึกษาที่เป็นเพศหญิง จะมีคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1) มีค่าเฉลี่ยอยู่ที่ 2.84 และคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) ค่าเฉลี่ยอยู่ที่ 2.83

ส่วนนักศึกษาที่ไม่สำเร็จการศึกษาที่เป็นเพศชาย จะมีคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1) มีค่าเฉลี่ยอยู่ที่ 2.01 และคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) ค่าเฉลี่ยอยู่ที่ 1.34

นักศึกษาที่ไม่สำเร็จการศึกษาที่เป็นเพศหญิง จะมีคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1) มีค่าเฉลี่ยอยู่ที่ 2.36 และคะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) ค่าเฉลี่ยอยู่ที่ 1.73

5.1.5 รายได้เฉลี่ยของผู้ปกครอง จะอยู่ 25,906.00 บาทต่อเดือน มีค่าเบี่ยงเบนมาตรฐานเท่ากับ 16,873.9449

5.1.6 คะแนนเฉลี่ยการประเมินรายวิชาที่นักศึกษาลงทะเบียนเรียนเท่ากับ 4.09 จากคะแนนเต็ม 5 มีค่าเบี่ยงเบนมาตรฐานเท่ากับ 0.4472

5.2 ประสิทธิภาพของการทำนายผลของโมเดล

การประเมินโมเดลในกลุ่ม Classification ซึ่งเป็นโมเดลสำหรับทำนายว่าสำเร็จการศึกษา (1) หรือไม่สำเร็จการศึกษา (0) ด้วยการใช้ Logistic Regression Model ได้ผลคะแนน Accuracy เท่ากับ 88.7% และ ได้ความแม่นยำในการทำนายของคนที่สำเร็จการศึกษาจริง และผลการทำนายได้ 88% และทำนายคนที่ไม่สำเร็จการศึกษาถูกต้อง 97%

ปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) ด้วย Logistic Regression Model พบว่าฟีเจอร์ที่มีผลต่อการทำนายสูงที่สุด คือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE), เพศ และ ผลคะแนนภาษาอังกฤษ จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา

ส่วนทำนายผลด้วยการใช้ Decision Tree Model จะได้ค่า Accuracy เท่ากับ 86.5% และได้ความแม่นยำในการทำนายของคนที่สำเร็จการศึกษาจริง และผลการทำนาย ได้ 88% และทำนายคนที่ไม่สำเร็จการศึกษาถูกต้อง 71% โดยในภาพ Decision Tree จะพบว่านักศึกษาที่มีผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 มากกว่า 1.46 ขึ้นไปจะสำเร็จการศึกษาทุกคน ส่วนนักศึกษาที่ได้คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 น้อยกว่า 1.46 จะมีโอกาสที่จะไม่สำเร็จการศึกษาสูง

การพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาที่รับด้วยวิธีรับตรงประจำปีการศึกษา 2561 ด้วย Decision Tree Model พบว่าปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) สูงที่สุด คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE) และ ระดับคะแนนเฉลี่ยการประเมินรายวิชาที่เรียน จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา

5.3 อภิปรายผลการวิจัย

ปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) ด้วย Logistic Regression Model พบว่าฟีเจอร์ที่มีผลต่อการทำนายสูงที่สุด คือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของวิชาที่ใช้ในการสอบแข่งขัน (SCORE), เพศ และ ผลคะแนนภาษาอังกฤษ จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา ในขณะที่การพยากรณ์ปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาที่รับด้วยวิธีรับตรงประจำปีการศึกษา 2561 ด้วย Decision Tree Model พบว่าปัจจัยหรือฟีเจอร์ที่มีผลต่อการทำนาย (Predictive Power) สูงที่สุด คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 2 ของการเรียนในชั้นปีที่ 1 (GPA2) รองลงมาคือ คะแนนผลการเรียนเฉลี่ยของภาคการศึกษาที่ 1 ของการเรียนในชั้นปีที่ 1 (GPA1), คะแนนรวมของการวิชาที่ใช้ในการสอบแข่งขัน (SCORE) และ ระดับคะแนนเฉลี่ยการประเมินรายวิชาที่เรียน จะมีผลต่อการทำนาย (Predictive Power) ลำดับถัดมา

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะเชิงนโยบาย

ควรจัดทำการธรรมาภิบาลข้อมูล (Data Governance) เพื่อให้ได้ข้อมูลที่มีคุณภาพและทันสมัย พร้อมใช้งาน เพื่อให้การทำนายด้วย Machine Learning มีความถูกต้องแม่นยำมากที่สุด โดยเฉพาะข้อมูลส่วนตัวของนักเรียน เช่น รายได้ของผู้ปกครอง

5.3.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

ในการศึกษาครั้งนี้มีการเลือกใช้ 2 โมเดลคือ Logistic Regression และ Decision Tree ในการวิเคราะห์ข้อมูลในภาพรวมทุกคณะ ควรทำการวิจัยต่อด้วยการแยกเป็นรายคณะ เพื่อให้สามารถ

นำไปใช้ประโยชน์ได้จริงในระดับคณะสาขา และทำการวิเคราะห์ข้อมูลในการรับเข้าศึกษาด้วยวิธีอื่นๆ เพื่อหาปัจจัยที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาที่รับด้วยวิธีรับต่างๆ ซึ่งจะมีเงื่อนไขที่แตกต่างกัน

เอกสารอ้างอิง

A Yunita et al (2021). Research Review on Big Data Usage for Learning Analytics and Educational Data Mining: A Way Forward to Develop an Intelligent Automation System. J. Phys.: Conf. Ser. 1898 012044

AIT Share. (2020, August 19). Course 1 [New] บทที่ 2 ส่วนที่ 1: 4 ประเภทการวิเคราะห์ Data Analytics [Video]. YouTube. https://www.youtube.com/watch?v=5BMxDbFi7y4

AIT Share. (2022a, August 24). รูปแบบจำลองจากการเรียนรู้ของเครื่อง (Machine Learning Models) [Video]. YouTube. https://www.youtube.com/watch?v=aY4fYKbTuxM

AIT Share. (2022b, August 25). อัลกอริทึม (Algorithms) [Video]. YouTube. https://www.youtube.com/watch?v=Y6OFdAMIftA

AIT Share. (2022c, August 26). การจำแนกประเภท (Classification) [Video]. YouTube. https://www.youtube.com/watch?v=ocbeuK3FZHY

Alkhalil, A., Abdallah, M. A., Alogali, A., & Aljaloud, A. (2021). Applying Big Data Analytics in Higher Education: A Systematic Mapping Study. International Journal of Information and Communication Technology Education (IJICTE), 17(3), 29-51. http://doi.org/10.4018/IJICTE.20210701.oa3

Bahga, A., & Madisetti, V. (2019). Big data science & analytics : a hands-on approach. Arshdeep Bahga & Vijay Madisetti.

Bahga, A., & Madisetti, V. (2019). Big data science & analytics : a hands-on approach. Arshdeep Bahga & Vijay Madisetti.

Cote C. (2021, November 02). WHAT ARE PRESCRIPTIVE ANALYTICS? 6 EXAMPLES. ค้นเมื่อวันที่ 26 กรกฎาคม 2565. https://online.hbs.edu/blog/post/prescriptive-analytics.

Dataiku. (2021). Machine Learning Basics: An Illustrated Guide for Non-Technical Readers. ค้นเมื่อวันที่ 24 กันยายน 2565 https://content.dataiku.com/ml-basics

Elliott, T. (2013, February 14). Analytics Moves to The Core. ค้นเมื่อวันที่ 26 กรกฎาคม 2565. https://timoelliott.com/blog/2013/02/gartnerbi-emea-2013-part-1-analytics-moves-to-the-core.html

Gavin M. (2019, July 16). BUSINESS ANALYTICS: WHAT IT IS & WHY IT'S IMPORTANT. ค้นเมื่อวันที่ 26 กรกฎาคม 2565. https://online.hbs.edu/blog/post/importance-of-business-analytics.

Gong, D. (February 23, 2022). Top 6 Machine Learning Algorithms for Classification. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501

GovTech. (2015, August 21). 7 Ways Smart Universities Use Data and Analytics. ค้นเมื่อวันที่ 25 ธันวาคม 2564, จาก https://www.govtech.com/education/7-ways-smart-universities-use-data-and-analytics.html

Grasso, C. (January 19, 2017). Machine Learning Explained: Algorithms Are Your Friend. ค้นเมื่อวันที่ 24 กันยายน 2565 https://blog.dataiku.com/machine-learning-explained-algorithms-are-your-friend

Joma Tech. (2018, June 23) What REALLY is Data Science? Told by a Data Scientist. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก https://www.youtube.com/watch?v=xC-c7E5PK0Y

Joshi, N. (2017, August 29). 4 Ways Big Data is Transforming the Education Sector. ค้นเมื่อวันที่ 24 ธันวาคม 2564, จาก https://www.allerin.com/blog/4-ways-big-data-is-transforming-the-education-sector.

K P Prakash and K. Selvakumari. (2021). Mathematical Modelling and Big-Data Analytics for Student Performance. J. Phys.: Conf. Ser. 1850 012017

Khan, S., & Alqahtani, S. (2020). Big Data Application and its Impact on Education. International Journal of Emerging Technologies in Learning (iJET), 15(17), pp. 36–46. https://doi.org/10.3991/ijet.v15i17.14459

Matthews, K. (2018, October 21). 6 Ways big data is improving higher education. ค้นเมื่อวันที่ 25 ธันวาคม 2564, จาก https://www.technobuffalo.com/6-ways-big-data-is-improving-higher-education

McCullum, N. (JUNE 29, 2020). How to Build and Train Linear and Logistic Regression ML Models in Python. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก https://www.freecodecamp.org/news/how-to-build-and-train-linear-and-logistic-regression-ml-models-in-python/

Meng, L. (2014). Application of Big Data in Higher Education. 2nd International Conference on Teaching and Computational Science (ICTCS 2014). https://www.atlantis-press.com/article/13336.pdf

Satangmongkol, K. (December 25, 2020). Hello World 03 – Logistic Regression. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก https://datarockie.com/blog/live-newyear-logistic-regression/

Satangmongkol, K. (March 21, 2021). อธิบาย 10 Metrics พื้นฐานสำหรับวัดผลโมเดล Machine Learning. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก https://datarockie.com/blog/top-ten-machine-learning-metrics/

Simplilearn. (2018, December 4). Data Science In 5 Minutes | Data Science For Beginners | What Is Data Science? YouTube. https://www.youtube.com/watch?v=X3paOmcrTjQ

Spear, E. (n.d.). 5 Big Benefits of Big Data Analytics in Education. ค้นเมื่อวันที่ 24 ธันวาคม 2564, จาก https://precisioncampus.com/blog/benefits-big-data-education/

Subramaniam, A. (2022, May 20). Big Data Analytics – Turning Insights Into Action. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก https://www.edureka.co/blog/big-data-analytics/

Tasmin, R., Muhammad, R. N., & Nor Aziati, ,A.H. (2020). Big data analytics applicability in higher learning educational system. IOP Conference Series.Materials Science and Engineering, 917(1) doi:http://dx.doi.org/10.1088/1757-899X/917/1/012064

Tuaha, S., Siddiqui, I.F., & Ali Arain, Q. (2019). Analyzing Students' Academic Performance through Educational Data Mining. 3C Tecnología_Glosas de innovación aplicadas a la pyme.

Viloria, A., Alexa, S. N., Hugo Hernández Palma, Núñez, W. N., & Núñez, L. N. (2020). Using big data to determine potential dropouts in higher education. Journal of Physics: Conference Series, 1432(1) doi:http://dx.doi.org/10.1088/1742-6596/1432/1/012077

Visitsora-at, P. (October 29, 2019). Metrics พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล Machine Learning. ค้นเมื่อวันที่ 24 มิถุนายน 2564, จาก

https://medium.com/@615162020027/metrics-พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล-machine-learning-c00fcc32fa30

ข้อมูลพื้นฐานเกี่ยวกับ Machine Learning (ต่อ): การสร้างแบบจำลอง Machine Learning ชุดแรกของคุณ. ค้นเมื่อวันที่ 24 กันยายน 256ถ, จาก https://content.dataiku.com/ml-basics-th

ธนพล ตั้งชูพงศ์. (2565). เอกสารประกอบการฝึกอบรม “หลักสูตรฝึกอบรมและปฏิบัติการเร่งรัดด้านวิทยาการข้อมูลและการประยุกต์” (Data Science and Applications) วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น https://www.canva.com/design/DAFDzTNWeGo/-s0Zs5Wm6KXwyRaZIKfYGg/view?utm_content=DAFDzTNWeGo&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink#23

วรรณา ญานวัฒนะ. (2551). **การเปรียบเทียบการพยากรณ์แผนการเรียนศึกษาต่อระดับมัธยมศึกษาตอนปลายด้วยโครงข่ายประสาทเทียมและการวิเคราะห์การถดถอยโลจิสติก กรณีศึกษา: โรงเรียนรัฐบาลแห่งหนึ่ง**. มหาวิทยาลัยธรรมศาสตร์.

ภาคผนวก

ประวัติผู้วิจัย

หัวหน้าโครงการวิจัย

1. ตำแหน่งในโครงการ หัวหน้าโครงการวิจัย
2. ชื่อ-ชื่อสกุล นายสุภาพ ไชยยา
   ตำแหน่งปัจจุบัน นักเทคโนโลยีสารสนเทศชำนาญการ
3. สังกัด/หมายเลขโทรศัพท์/โทรสาร/e-mail
   งานรับเข้าและการตลาด สำนักบริหารและพัฒนาวิชาการ
   โทร 084-600-4391 E-mail chaisupap@kku.ac.th

4. ประวัติการศึกษา (ปี พ.ศ.ที่จบ ระดับปริญญา คุณวุฒิ สาขาวิชา สถานศึกษา ประเทศ)

พ.ศ.	ระดับ	คุณวุฒิ	สาขาวิชา	สถานศึกษา
2532	ปริญญาตรี	วิทยาศาสตร์บัณฑิต	สัตวศาสตร์	มหาวิทยาลัยขอนแก่น
2560	ปริญญาโท	รัฐประศาสนศาสตรมหาบัณฑิต	การปกครองท้องถิ่น	มหาวิทยาลัยขอนแก่น

5. สาขาวิชาการที่มีความชำนาญ/เชี่ยวชาญ (ถ้ามี)
   การวิเคราะห์ข้อมูลทางด้วยโปรแกรม Power BI
6. ประสบการณ์ที่เกี่ยวข้องกับงานวิจัยทั้งในและต่างประเทศ (งานวิจัยที่ทำเสร็จแล้ว : ชื่อเรื่อง ปีที่พิมพ์ และสถานที่ในการวิจัย, งานวิจัยที่กำลังทำ : ชื่อเรื่อง แหล่งทุน สถานะโครงการ)
   ความพึงพอใจของผู้สมัครต่อการสมัครเข้าศึกษาในมหาวิทยาลัยขอนแก่นในระดับปริญญาตรี โดยวิธีรับตรงผ่านทางระบบ Internet ประจำปีการศึกษา 2556 วารสารวิจัยสถาบัน มข. KKU Institutional Research Journal ปีที่ 2 ฉบับพิเศษ ประจำเดือน มิถุนายน พ.ศ. 2557 Vol.2 (Supplement) June. 2014 ISSN 2350-9163

ผู้ร่วมโครงการวิจัย

1. ตำแหน่งในโครงการ ผู้ร่วมโครงการวิจัย
2. ชื่อ-ชื่อสกุล นายวิทวัส เหล่ามะลอ
   ตำแหน่งปัจจุบัน นักวิชาการแผนและสารสนเทศ ระดับปฏิบัติการ
3. สังกัด/หมายเลขโทรศัพท์/โทรสาร/e-mail
   งานรับเข้าและการตลาด สำนักบริหารและพัฒนาวิชาการ
   โทร 095-253-1428 E-mail witthla@kku.ac.th

4. ประวัติการศึกษา (ปี พ.ศ.ที่จบ ระดับปริญญา คุณวุฒิ สาขาวิชา สถานศึกษา ประเทศ)

พ.ศ.	ระดับ	คุณวุฒิ	สาขาวิชา	สถานศึกษา
2553	ปริญญาตรี	วิทยาบัณฑิต	คณิตศาสตร์ประยุกต์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าลาดกระบัง
2557	ปริญญาโท	วิทยาศาสตร์มหาบัณฑิต	สถิติประยุกต์	มหาวิทยาลัยธรรมศาสตร์

5. สาขาวิชาการที่มีความชำนาญ/เชี่ยวชาญ (ถ้ามี)
   การวิเคราะห์ข้อมูลด้วยโปรแกรมสถิติ SPSS
6. ประสบการณ์ที่เกี่ยวข้องกับงานวิจัยทั้งในและต่างประเทศ (งานวิจัยที่ทำเสร็จแล้ว : ชื่อเรื่อง ปีที่พิมพ์ และสถานที่ในการวิจัย, งานวิจัยที่กำลังทำ : ชื่อเรื่อง แหล่งทุน สถานะโครงการ)
   ปัจจัยที่มีผลต่อการตัดสินใจเข้าศึกษาต่อระดับปริญญาตรีของนักศึกษามหาวิทยาลัยขอนแก่น ประจำปีการศึกษา 2562 โดยผ่านการคัดเลือกด้วยระบบ TCAS