

**ระบบวิเคราะห์ และ พยากรณ์ สำหรับการบริหารหลักสูตรวิศวกรรม  
คอมพิวเตอร์**

**Analytics and Prediction System for CE Curriculum administrators**

**ณิกานต์ สุขุมจิตพิทยภัท**

**นรวิชญ์ อยู่บัว**

**ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต**

**สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์**

**สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง**

**ปีการศึกษา 2565**

ปริญญานิพนธ์ปี การศึกษา 2564

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ระบบวิเคราะห์ และ พยากรณ์ สำหรับการบริหารหลักสูตรวิศวกรรมคอมพิวเตอร์

Analytics and Prediction System for CE Curriculum administrators

ผู้จัดทำ

- |                  |                   |                       |
|------------------|-------------------|-----------------------|
| 1. นางสาวณิกานต์ | สุขุมจิตพิทยโยทัย | รหัสนักศึกษา 62010299 |
| 2. นายนรวิษฐ์    | อยู่บัว           | รหัสนักศึกษา 62010465 |

อาจารย์ที่ปรึกษา

(ผศ. ดร. ธนัญชัย ศรีภาค)

# ระบบวิเคราะห์ และ พยากรณ์ สำหรับการบริหารหลักสูตรวิศวกรรม คอมพิวเตอร์

นางสาวณิกานต์	สุขุมจิตพิทยุทธ	62010299
นายณรวิชญ์	อยู่บัว	62010465
ผศ. ดร. ธนัญชัย	ตรีภาค	อาจารย์ที่ปรึกษา
ปีการศึกษา 2565		

## บทคัดย่อ

โครงการนี้จัดทำขึ้นเพื่อ พัฒนาระบบประมวลผลข้อมูลผลการเรียนของนักศึกษาในอดีต ข้อมูลของรายวิชาต่างๆ และข้อมูลจากแบบสำรวจการมีงานทำของบัณฑิต เพื่อนำเสนอข้อมูลสถิติต่างๆ วิเคราะห์ข้อมูลผลการผลิตบัณฑิตเพื่อให้ได้ผลลัพธ์ว่าที่ผ่านมาหลักสูตรสามารถผลิตบัณฑิตกลุ่มใดได้บ้าง มีจำนวนมากน้อยเพียงใด สามารถพยากรณ์ว่าในอนาคตหลักสูตรสามารถผลิตบัณฑิตกลุ่มใดได้เป็นจำนวนเท่าใด เพื่อเป็นประโยชน์และอำนวยความสะดวกให้กรรมการหลักสูตรในการวางแผนการบริหารหลักสูตรในอนาคต และแสดงเป็นแผนภาพกราฟิกในการอำนวยความสะดวกให้หน่วยงานภายนอกได้รับทราบว่าหลักสูตรปัจจุบันของสถาบันสามารถผลิตบุคลากรที่มีความชำนาญด้านใดได้บ้าง

# **Analytics and Prediction System for CE Curriculum**

## **administrators**

Ms. Nichakan Sukhumjitpitayotai 62010299

Mr. Narawich Youbua 62010465

Mr. Thanunchai Threepak Advisor

Academic Year 2022

## **Abstract**

## กิตติกรรมประกาศ

โครงการในภาคการศึกษานี้สำเร็จลุล่วงได้ด้วยดีจากความช่วยเหลือจากหลากหลายบุคคล โครงการในภาคการศึกษานี้จะผ่านไปไม่ได้หากปราศจากความช่วยเหลือจากบุคคลเหล่านี้ขอขอบคุณอาจารย์ที่ปรึกษา ผศ. ดร. ธนัญชัย ตรีภาค ที่ให้ความช่วยเหลือในเรื่องต่าง ๆ ไม่ว่าจะเป็นการให้คำแนะนำถึงแนวทางการทำงานที่ดี การให้คำปรึกษาเพื่อหาทางออกเมื่อพบเจอกับปัญหา รวมถึงให้ความรู้เกี่ยวกับตัวงานทำให้งานต่าง ๆ เมื่อเจอปัญหา ก็สามารถผ่านไปได้อย่างดี

ขอขอบคุณคณาจารย์ในภาควิชาวิศวกรรมคอมพิวเตอร์ ที่ประสาทวิชาการความรู้มาตลอด 4 ปี ซึ่งความรู้หลาย ๆ แขนงก็ถูกใช้เป็นพื้นฐาน และเป็นส่วนหนึ่งของโครงการนี้

ขอขอบคุณเพื่อน ๆ วิศวกรรมคอมพิวเตอร์ที่ให้คำปรึกษา และแลกเปลี่ยนความรู้ซึ่งกันและกัน รวมถึงการรับฟังปัญหา

สุดท้ายนี้ขอขอบคุณบิดา มารดาและครอบครัว ที่เลี้ยงดูอบรมสั่งสอนและให้ความรู้คุณธรรม จริยธรรม และให้การสนับสนุนด้านการศึกษาจนได้มีโอกาสมาทำโครงการนี้

ณิชากรณต์ สุขุมจิตพิทโยทัย

นรวิชญ์ อยู่บัว

# สารบัญ

หน้า

บทคัดย่อภาษาไทย

บทคัดย่อภาษาอังกฤษ

กิตติกรรมประกาศ

สารบัญ

สารบัญตาราง

สารบัญภาพ

บทที่ 1 บทนำ

1.1 ความเป็นมาของปัญหา

1.2 วัตถุประสงค์

1.3 ประโยชน์ของโครงการ

1.4 ขอบจำกัดของโครงการ

1.5 แผนการดำเนินงาน

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.2 เครื่องมือที่เกี่ยวข้อง

2.3 งานวิจัยที่เกี่ยวข้อง

บทที่ 3 การออกแบบ

บทที่ 4 ผลการดำเนินงาน

## สารบัญ(ต่อ)

บทที่ 5 สรุป

5.1 บทสรุป

5.2 ปัญหาและอุปสรรคที่พบ

5.3 แนวทางในการพัฒนาต่อ

เอกสารอ้างอิง

ภาคผนวก

# สารบัญตาราง

ตาราง

หน้า



# สารบัญรูป

รูป

หน้า

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาของปัญหา

Data Analytics เป็นการวิเคราะห์ข้อมูลที่มีอยู่ตั้งแต่อดีตจนถึงปัจจุบัน ในกรณีที่ข้อมูลเพียงพอและเหมาะสมจะสามารถนำมาคาดการณ์แนวโน้ม ทำนายอนาคตที่เป็นประโยชน์ พยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้นโดยใช้ข้อมูลในอดีตกับแบบจำลองทางสถิติรวมถึงการให้คำแนะนำทางเลือกต่าง ๆ และผลของแต่ละทางเลือก

จากปัญหาที่ทางผู้จัดทำเล็งเห็นความสำคัญคือการนำข้อมูลผลการเรียนของนักศึกษาในอดีตมาใช้ประโยชน์ในการบริหารหลักสูตร และ นำมาวิเคราะห์ผลเพื่อช่วยในการวางแผนการเรียนของนักศึกษา ซึ่งการวางแผนในการเรียนของหลักสูตรจะสามารถช่วยอาจารย์และบุคลากรที่เกี่ยวข้องกับการศึกษาในด้านของการบริหารหลักสูตร เพื่อวางแผนการเพิ่มหรือลดจำนวนผู้เรียนในรายวิชาต่าง ๆ ซึ่งส่งผลต่อการผลิตบัณฑิตด้านต่าง ๆ ได้

ดังนั้นผู้จัดทำจึงได้เห็นถึงความสำคัญการประเมินสถานะขอหลักสูตร ของระบบแนะนำการวางแผนการคาดการณ์จากการใช้ความรู้ทางด้าน Data Analytics, Prediction และ Recommendation โดยใช้ข้อมูลผลการเรียนของนักศึกษาในอดีต เพื่อพัฒนาระบบช่วยเหลือ และตอบโจทย์ให้แก่นักศึกษาและบุคลากรทางการศึกษาหรือบุคคลที่เกี่ยวข้องได้

### 1.2 วัตถุประสงค์

- 1) เพื่อนำข้อมูลของผลการเรียนของนักศึกษาในอดีตและข้อมูลจากแบบสำรวจการมีงานทำของบัณฑิตมาใช้ ในการวางแผนการเรียนหรือประเมินอาชีพในอนาคตของนักศึกษาได้
- 2) ประมวลผลข้อมูลผลการเรียนของนักศึกษาในอดีต และข้อมูลจากแบบสำรวจการมีงานทำของบัณฑิต และทำแผนภาพกราฟิกเพื่อนำเสนอข้อมูล อำนวยความสะดวกให้กรรมการหลักสูตรในการวางแผนการ ทำงาน
- 3) เพื่อนำข้อมูลผลการเรียนของนักศึกษาในอดีต มาพัฒนาเป็นระบบแนะนำและวางแผนการเรียนตัวของ นักศึกษาได้

- 4) เพื่อนำข้อมูลการพยากรณ์อาชีพในอนาคตของนักศึกษาในสถาบันมาแสดงเป็นแผนภาพกราฟิกในการอำนวยความสะดวกให้หน่วยงานภายนอกได้รับทราบว่าหลักสูตรปัจจุบันของสถาบันสามารถผลิตบุคลากรที่มีความชำนาญด้านใดได้บ้าง

### 1.3 ประโยชน์ของโครงการ

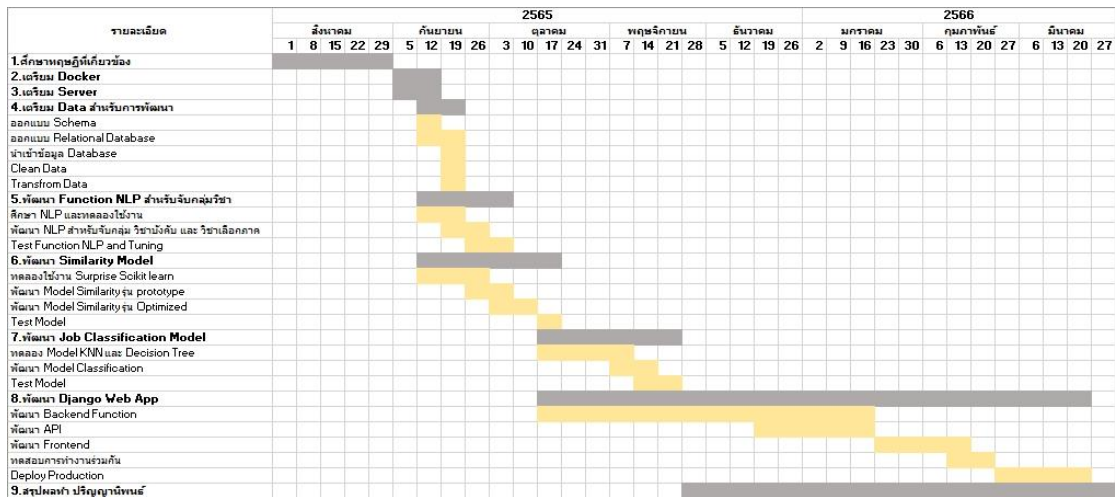
- 1) ได้รวบรวมข้อมูลผลการเรียนของนักศึกษาและข้อมูลแบบสำรวจการทำงานของบัณฑิต แล้วนำมาวิเคราะห์และนำเสนอข้อมูลที่เป็นประโยชน์ในการบริหารหลักสูตรของกรรมการหลักสูตร
- 2) มีระบบที่สามารถแนะนำ วางแผน และประเมินอาชีพในอนาคตจากผลการเรียนของนักศึกษา

### 1.4 ข้อจำกัดของโครงการ

- 1) ข้อมูลผลการเรียนในอดีตย้อนหลังมีเพียง 2 ปี
- 2) ข้อมูลผลการเรียนในอดีตจะได้จากสำนักทะเบียนและประมวลผล โดยกรรมการหลักสูตรจะเป็นผู้ร้องขอข้อมูลดังกล่าวและนำเข้าระบบ
- 3) การทำนายต่าง ๆ จะใช้ข้อมูลเพียง 2 แหล่งคือข้อมูลผลการเรียนของนักศึกษาจากสำนักทะเบียนและประมวลผล และแบบสอบถามการปฏิบัติงานของบัณฑิตเท่านั้น

## 1.5 แผนการดำเนินงาน

แผนการดำเนินงานในการพัฒนาโครงการตลอดระยะเวลา 2 ภาคการศึกษา ตั้งแต่เดือนสิงหาคม พ.ศ. 2565 - มีนาคม พ.ศ. 2566 แสดงดังรูป 1.1



รูป 0.1 แผนการดำเนินงาน

## บทที่ 2

# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

## 2.1 ทฤษฎีที่เกี่ยวข้อง

### 2.1.1 Classification and Prediction

Classification and Prediction คือการจำแนกประเภทของข้อมูล โดยจะนำมาใช้ในการวิเคราะห์ข้อมูล ซึ่งกระบวนการดังกล่าวสามารถแบ่งได้เป็น 2 ส่วน

- 1) Training Data คือการนำข้อมูลมาที่ได้มาทำการเรียนรู้ให้กับคอมพิวเตอร์เพื่อนำไปสร้างเป็นโมเดลแบบจำลองและวัดประสิทธิภาพของโมเดลแบบจำลองนั้น โดยจะทำการสร้างโมเดลซึ่งจะมีด้วยกันหลายวิธี เช่น Decision Tree, Naive Bayes, K Nearest Neighbors และ Neural Network เป็นต้น
- 2) Predict คือการนำข้อมูลใหม่ที่รับมานำเข้าโมเดลแบบจำลองที่เป็นผลลัพธ์จากการผ่านกระบวนการ Training Data ไปทำการคำนวณหรือพยากรณ์

ประเภทของปัญหาในด้าน Classification

- 1) Binary classification (การจำแนกแบบไบนารี)  
เปรียบเทียบให้ดีที่สุดคือ ตัวแปรที่อยู่ในรูปแบบสองหมวดหมู่ เช่น ผลลัพธ์แบบใช่ หรือ ไม่ใช่ ตก หรือ ผ่าน หากเปรียบเทียบในรูปแบบของตัวเลขก็คือ 0 กับ 1 อัลกอริทึมที่ใช้คู่กับการจำแนกแบบไบนารี จะมีดังนี้ k-Nearest Neighbors Decision Trees หรือ Naive Bayes
- 2) Multi-Class Classification (การจำแนกประเภทหลายคลาส)  
ในการจำแนกรูปแบบนี้จะต่างกับการจำแนกแบบไบนารี โดยจะมีหมวดหมู่มากกว่าสอง ตัวอย่างของการจำแนกประเภทนี้ เช่น รูปภาพที่มีองค์ประกอบคล้ายคลึงกับรูปภาพที่อยู่ในฐานข้อมูลเพื่อค้นหาคำศัพท์ที่คาดว่าจะพิมพ์ใน predictive keyboard โดยผลลัพธ์ที่อาจเกิดนั้นจะมีได้มากกว่า 2 หมวดหมู่

อัลกอริทึมที่ใช้คู่ไปกับการจำแนกประเภทนี้สามารถใช้อัลกอริทึมคล้ายกับการจำแนกแบบไบนารีได้

3) Multi-Label Classification (การจำแนกประเภทหลายเลเบล)

เปรียบให้เข้าใจง่ายโดยการยกตัวอย่างเช่น รูปภาพรูปหนึ่งสามารถมีรูปดอกไม้ ท้องฟ้าก้อนเมฆได้ แต่รูปภาพรูปนั้นจะจัดว่าเป็นหมวดหมู่รูปภาพรูปถ่าย หรือรูปเสีย Multi-Label Classification ก็คือการทำเลเบลให้กับชุดข้อมูล หรือการตัดสินใจให้รูปนั้น ๆ ว่ามีดอกไม้หรือเปล่ามีก้อนเมฆหรือไม่ ส่วน Multi-Class Classification จะจำแนกว่ารูปนั้นเป็นรูปที่เกิดจากการวาดหรือรูปที่เกิดจากการถ่ายหรือรูปเสีย

4) Imbalanced Classification (การจำแนกแบบข้อมูลไม่เท่าเทียม)

คือปัญหาที่เกิดจากข้อมูลที่มีไม่เท่าเทียมกัน (Imbalanced dataset) ตัวอย่างเช่นข้อมูลของการทุจริตโดยข้อมูลส่วนใหญ่ย่อมเป็นข้อมูลที่จัดว่า “ไม่ทุจริต” และจะมีเปอร์เซ็นต์น้อยที่จัดว่าเป็น “ทุจริต” เป็นต้น โดยจะเปรียบโดยง่ายคือกรณีที่ชุดข้อมูลมีการแยกประเภทกันแต่จำนวนของประเภทนั้นมีอัตราส่วนของข้อมูลที่ห่างกันค่อนข้างมาก

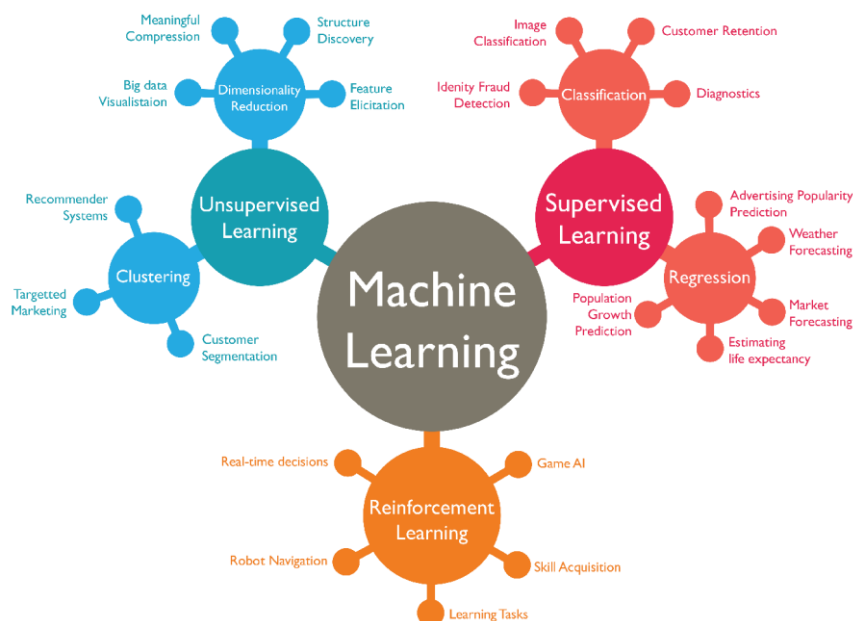
### 2.1.2 Machine Learning

Machine Learning คือ การทำให้ระบบของคอมพิวเตอร์นั้นสามารถเรียนรู้ได้ด้วยตนเอง โดยจะใช้ข้อมูล ด้วยวิธีการใส่ข้อมูลและผลลัพธ์เข้าไป เพื่อให้โปรแกรมนำผลลัพธ์นั้นไปประมวลผลและพยากรณ์ Output และ Input ของข้อมูลใหม่ โดยแบ่ง Machine Learning ออกได้เป็น 3 ประเภท คือ

- 1) Supervised Learning คือการเรียนรู้ที่เครื่องจักร หรือคอมพิวเตอร์นั้น จำเป็นต้องอาศัยข้อมูลใน การฝึกฝน เปรียบเสมือนกับการเรียนการสอนของเด็ก ซึ่งจำเป็นอย่างยิ่งที่จะต้องอาศัยชุดของข้อมูล ซึ่งประกอบไปด้วยชุดของข้อมูล และชุดของผลลัพธ์ของข้อมูลที่ต้องการจะนำมาให้ เครื่องจักรหรือคอมพิวเตอร์ในการเรียนรู้
- 2) Unsupervised Learning เป็นการเรียนรู้ที่ให้เครื่องจักรหรือคอมพิวเตอร์นั้น สามารถเรียนรู้ได้ด้วยตนเอง โดยไม่จำเป็นต้องมีค่าเป้าหมายของแต่ละชุดข้อมูล ซึ่งวิธีการนี้คือการที่มนุษย์นั้นจะเป็นผู้ใส่ชุดข้อมูล และกำหนดสิ่งที่

ต้องการจากชุดข้อมูลเหล่านั้น โดยให้เครื่องจักรหรือคอมพิวเตอร์วิเคราะห์  
จากการจำแนกและทำการสร้างแบบแผนจากข้อมูลที่ได้รับมา

- 3) Reinforcement Learning เป็นการเรียนรู้สิ่งต่าง ๆ ผ่านจากการลองผิดลองถูก  
ภายใต้แนวคิดที่ว่าจะเลือกกระทำสิ่งใดที่ทำให้ได้ผลลัพธ์มากที่สุด โดยจะทำการ  
เรียนรู้จากการลองผิดลองถูกในสถานการณ์ในอดีตหรือระบบจำลอง  
และพยายามที่จะพัฒนาระบบการตัดสินใจของตัวเองให้ดีขึ้นอย่างต่อเนื่อง  
โดยที่อาจจะสามารถพัฒนาด้วยการพยายามสร้างแบบจำลองสถานการณ์ต่าง  
ๆ ขึ้นมา



รูป 0.1 ประเภทของ Machine Learning

(ที่มา : [medium.com](https://medium.com/), 2018)

### 2.1.3 Extract-Transform-Load (ETL)

Extract-Transform-Load คือ กระบวนการ กระบวนการหนึ่งซึ่งอยู่ในระบบของ  
Data Warehouse ซึ่งเป็นระบบที่ออกแบบมาเพื่อที่จะสามารถดึงข้อมูลออกมาจากหลายแหล่ง  
โดยจะนำกระบวนการตรวจสอบคุณภาพของชุดข้อมูลมาประยุกต์ร่วมใช้ ซึ่งมีการเชื่อมโยงและ  
ปรับชุดของข้อมูลให้เป็นไปในรูปแบบเดียวกันทั้งหมดเพื่อให้ ชุดของข้อมูลจากหลากหลาย  
แหล่งสามารถใช้งานร่วมกันได้ และทำการส่งมอบ

- 1) Extract เป็นกระบวนการเริ่มต้นของระบบที่ดึงข้อมูลจากแหล่งของข้อมูล จะประกอบด้วยข้อมูลจากหลากหลายแหล่งที่มา ข้อมูลที่อยู่ต่างที่กันนั้นอาจจะอยู่ในรูปแบบที่แตกต่างกัน ยกตัวอย่างเช่น อาจจะอยู่ในรูปแบบของฐานข้อมูลคนละชนิด หรือ ไม่ใช่ฐานข้อมูลแท้จริงซึ่งอาจจะเป็นระบบไฟล์ข้อมูลธรรมดา
- 2) Transforming ขั้นตอนการแปลงรูปแบบของข้อมูลนี้จะมีการใช้กฎหรือฟังก์ชัน (Function) มากมายเพื่อที่จะแปลงข้อมูลให้อยู่ในรูปแบบตามที่ต้องการก่อนที่จะนำข้อมูลเหล่านั้นเข้าไปยังปลายทาง ข้อมูลจากต้นทางนั้นบางแหล่งข้อมูลมีความจำเป็นน้อยมากหรือแทบจะไม่ต้องการ การแปลงข้อมูลเลย แต่ในบางแหล่งอาจจะต้องการกระบวนการที่ซับซ้อนในการแปลงข้อมูล ซึ่งจะกินทรัพยากรของระบบที่ใช้และเวลาในการประมวลผลของระบบ ซึ่งความซับซ้อนของข้อมูลจะขึ้นอยู่กับความต้องการของเชิงธุรกิจ หรือ เป้าหมายของการนำข้อมูลไปใช้งาน โดยจะมีกระบวนการตัวอย่างต่อไปนี้
  - 1) Selection คือ การเลือก Column ที่ต้องการที่จะนำไปใช้งานหรือเก็บลงฐานข้อมูล ยกตัวอย่าง เช่น ถ้าต้นทางของข้อมูลมีอยู่ด้วยกัน 3 Column หรือ 3 attributes เช่น enroll\_num, age และ salary จะมีการแปลงข้อมูลเกิดขึ้นและ เลือกที่จะไม่มีการแปลงข้อมูลหากพบว่า record นั้นมีค่าของข้อมูล column salary เป็นค่าว่าง
  - 2) Translation คือ การแปลงข้อมูล ตัวอย่างเช่น หากข้อมูลต้นทางนั้นมีการเก็บข้อมูลของเพศโดยให้ 1 เป็นเพศชาย และ 2 เป็นเพศหญิง จะต้องมีการแปลงจากชุดตัวเลขที่กำหนดก่อนหน้านี้ให้ 1 = Male และ 2 = Female กระบวนการนี้เรียกว่า data cleaning หรือ กระบวนการทำความสะอาดข้อมูล
  - 3) Encoding free form ยกตัวอย่างเช่นการ mapping จาก “Male” ไปเป็น “1” และ “Mr” ไปเป็น “M”
  - 4) Filtering คือ กระบวนการกรองเฉพาะข้อมูลที่กำหนด
  - 5) Sorting คือ กระบวนการเรียงข้อมูลที่ต้องการ
  - 6) Joining คือ กระบวนการเชื่อมโยงข้อมูลระหว่างตารางข้อมูล



- 7) Aggregation คือ กระบวนการรวบรวม และ สรุปชุดข้อมูล ยกตัวอย่าง เช่น การรวมยอด (summarize) ข้อมูลจากหลาย ๆ ระเบียบจนได้มาเป็น ยอดขายรวม เป็นต้น
- 8) Transposing or pivoting คือการสลับทิศทางการตำแหน่งของการแสดงข้อมูล เช่นการย้ายระเบียบไปเป็น Column หรือ ย้าย Column มาเป็น ระเบียบ เพื่อให้ง่ายต่อการนำข้อมูลไปใช้
- 3) Loading กระบวนการโหลดข้อมูลเข้า โดยทั่วไปจะนำข้อมูลเข้าไปในระบบ Data Warehouse ทั้งนี้ขึ้นอยู่กับความต้องการขององค์กร หรือ ธุรกิจว่าจะให้ข้อมูลไหลไปในทิศทางใด บางองค์กร หรือ บางงานจะมีการสะสมของข้อมูล ความถี่ของการนำข้อมูลเข้าสู่ระบบ อาจจะมีการล้างข้อมูลแล้วทับข้อมูลใหม่ โดยทั่วไปแล้วข้อมูลของ Data Warehouse จะมีการใช้กันปีต่อปี เมื่อขึ้นปีใหม่แล้วจะมีการล้างข้อมูลของปีเก่า และ เก็บไว้ในระบบข้อมูลสำรอง เนื่องจากว่ากระบวนการนำข้อมูลเข้าจะต้องปฏิบัติตามพันธกิจกับฐานข้อมูล (Database) ดังนั้น จะต้องมีการมีประเด็นเรื่อง ของ Database Constraints, Referential Integrity, Database Trigger เข้ามาเกี่ยวข้องด้วยในกระบวนการนำข้อมูลเข้า ซึ่งสิ่งเหล่านี้รวม ๆ แล้วเรียกว่า กระบวนการควบคุมคุณภาพของข้อมูล (Data Quality performance of E-T-L process)

#### 2.1.4 Natural Language Processing (NLP)

Natural Language Processing (NLP) เป็นเครื่องมือที่ให้คอมพิวเตอร์เข้าใจภาษาของมนุษย์ที่มีความซับซ้อน เป็นศาสตร์หนึ่งที่สำคัญทางด้าน Machine Learning โดยเป็นสาขาวิชาหนึ่งที่ประกอบด้วยองค์ความรู้จากหลากหลายแขนง อาทิ ภาษาศาสตร์ (Linguistics) วิทยาการคอมพิวเตอร์ (Computer Science) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) รวมไปถึงสถิติ (Statistics) โดย NLP มีมาตั้งแต่ช่วงกลางศตวรรษที่ 19 และมีการพัฒนามาเรื่อย ๆ จนถึงปัจจุบัน โดยแบ่งออกเป็น 3 ยุค ดังนี้

- 1) ยุค Rule-based Method (ช่วง ค.ศ.1950-1990)

ในยุคแรกของ NLP มีการใช้งานตามกฎ (Rule-based Method) โดยนักภาษาศาสตร์ผู้มีความเชี่ยวชาญโครงสร้างของภาษาที่สนใจ จะเป็นผู้เขียน

กฎต่าง ๆ ขึ้นมาเพื่อให้คอมพิวเตอร์สามารถคำนวณข้อความของโจทย์ต่างๆ ได้

2) ยุค Machine Learning (ช่วง ค.ศ.1990-2010)

ในยุคนี้ พบว่ามีการเขียนกฎด้วยมือไม่สามารถตอบโจทย์ที่มีความซับซ้อนได้ จึงมีสิ่งที่ได้มาทดแทนในยุคนี้คือ ความสามารถของเครื่องคอมพิวเตอร์ รวมถึงความรู้ทางด้านสถิติ และ Machine Learning ซึ่งได้ถูกนำมาพัฒนาเพื่อใช้ในการทำงานด้าน NLP โดยมีการนำเข้าข้อมูลเพื่อให้คอมพิวเตอร์สามารถเรียนรู้ด้วยตนเองแทนการใช้ผู้เชี่ยวชาญทางด้านภาษา

3) ยุค Deep Learning (ช่วง ค.ศ.2010-ปัจจุบัน)

ในยุคปัจจุบัน ด้วยพลังการคำนวณของคอมพิวเตอร์ที่มีการพัฒนาสูงขึ้นอย่างต่อเนื่อง ทำให้เทคโนโลยีที่มีความซับซ้อนสูงอย่าง การเรียนรู้เชิงลึก (Deep Learning) ถูกนำมาใช้งานแทนที่ Machine Learning ซึ่งใช้ความรู้ทางด้านสถิติแบบดั้งเดิมอย่างแพร่หลายมากขึ้น รวมถึงในงานด้าน NLP ด้วยเช่นกัน อาทิ การสร้างแบบจำลองทางภาษา (Language Model) และการวิเคราะห์โครงสร้างของข้อความ (Parsing)

ตัวอย่างการประยุกต์ใช้ NLP ในด้านต่าง ๆ

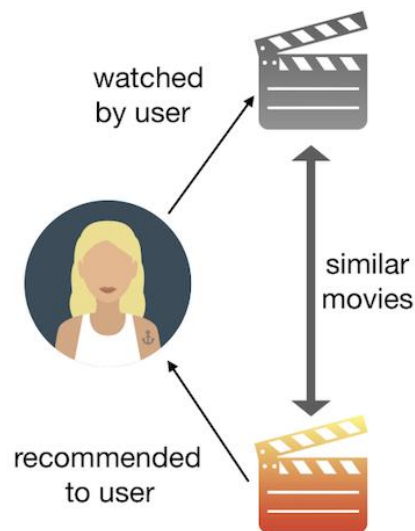
- 1) ด้านการทำงานวิจัย การวิจัยมีแหล่งของข้อมูลทางภาษานานาชนิด ซึ่งทำให้ NLP สามารถเข้ามามีบทบาทได้อย่างหลากหลาย ตัวอย่างเช่น การใช้ Topic Model ในการจัดหมวดหมู่บทความ
- 2) ด้านพาณิชย์อิเล็กทรอนิกส์ การซื้อของผ่านช่องทางออนไลน์ เข้ามามีบทบาทสำคัญเป็นอย่างมากในระบบเศรษฐกิจ ซึ่งทำให้เกิดปริมาณธุรกรรมขนาดใหญ่ ไม่ว่าจะเป็น คำอธิบายสินค้าและบริการ การแสดงความคิดเห็นของผู้บริโภค รวมถึงการสนทนากันระหว่างผู้ซื้อและผู้ขายผ่านทางช่องทาง
- 3) ด้านการแพทย์ ข้อมูลทางการแพทย์มีการบันทึกข้อมูลด้วยข้อความ ตัวอย่างเช่น บทสนทนาระหว่างแพทย์และผู้ป่วย การวินิจฉัยโรคโดยแพทย์ และประวัติการรักษาของผู้ป่วย
- 4) ด้านกฎหมาย สำหรับงานด้าน มีข้อมูลทางด้านภาษาที่แตกต่างและหลากหลาย เช่นเดียวกัน เช่น ประมวลกฎหมายต่าง ๆ คำร้องต่อศาล คำให้การของคู่ความ และคำพิพากษาของศาล ซึ่งสามารถประยุกต์ใช้

เครื่องมือ NLP ได้ในหลายมิติไม่ว่าจะเป็นการใช้ PoS Tagging และ NER เพื่อช่วยในการตีความประมวลผลกฎหมาย

### 2.1.5 Recommendation System

Recommendation System เป็นระบบที่จะทำการแนะนำสิ่ง (item) ที่ “เหมาะสม” ให้แก่ผู้ใช้ โดย item เป็นได้ตั้งแต่ ข้าว เนื้อหา เพลง course เรียน ไปจนถึงสินค้าที่ขายในร้าน online โดยสามารถแนะนำสิ่งที่ผู้ใช้สนใจได้ผ่านโมเดลที่ส่วนใหญ่มักจะถูกใช้กันมีอยู่ด้วยกันสามประเภท ได้แก่

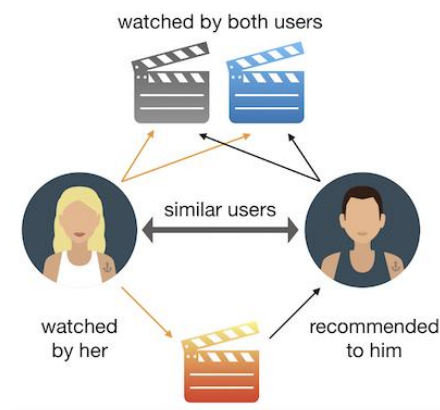
- 1) Content-based Filtering เป็นรูปแบบของโมเดลที่จะแนะนำลักษณะของตัวบริการหรือสินค้าเป็นตัวตั้งและทำการแนะนำสิ่งคำที่มีลักษณะที่คล้ายกัน



รูป 2.6 รูปแบบของ Content-based Filtering

(ที่มา : [towardsdatascience.com](https://towardsdatascience.com/), 2018)

- 2) Collaborative Filtering เป็นรูปแบบโมเดลที่เรียนรู้จากพฤติกรรมของผู้ใช้กับผู้ใช้คนอื่น ๆ ที่คล้ายคลึงกัน
  - 1) Memory-based เป็นการดูข้อมูลแล้วหาความสัมพันธ์ ระหว่างผู้ใช้หรือสินค้าจากข้อมูลโดยตรง



รูป 2.6 รูปแบบของ Memory-based

(ที่มา : [towardsdatascience.com](https://towardsdatascience.com), 2018)

- 2) Model-based ใช้เทคนิคของ machine learning เพื่อหา user embedding และ item embedding มาทำการทำนาย rating ที่ผู้ใช้จะให้กับสินค้า หรือ relevance score
- 3) Hybrid ใช้หลาย ๆ วิธีการมารวมกัน Hybrid system เป็นการมักรวมทั้งสองอัลกอริทึมของ Model-based และ Memory-based เอาไว้เพื่อทำให้ระบบการแนะนำสมบูรณ์ขึ้น ซึ่งระบบนี้ถูกนำไปใช้ในปัจจุบันมากที่สุดแทบจะทุกแพลตฟอร์มใหญ่ที่มีการแนะนำสินค้าและบริการ
- 3) Hybrid system เป็นการมักรวมทั้งสองระหว่าง Content-based Filtering และ Collaborative Filtering เพื่อทำให้ระบบการแนะนำสมบูรณ์ขึ้น

## 2.2 เครื่องมือที่เกี่ยวข้อง

### 2.2.1 Docker

Docker เป็นเครื่องมือแบบ open-source ที่ช่วยจำลองสภาพแวดล้อม ในการรัน service หรือ server โดยการสร้าง container เพื่อจัดการกับ library ต่างๆ และยังช่วยจัดการในเรื่องของ version control เพื่อให้ง่ายต่อการจัดการกับปัญหาต่างๆ ที่เกิดขึ้นของค์ประกอบต่างๆ ของ Docker

#### 1) Docker image

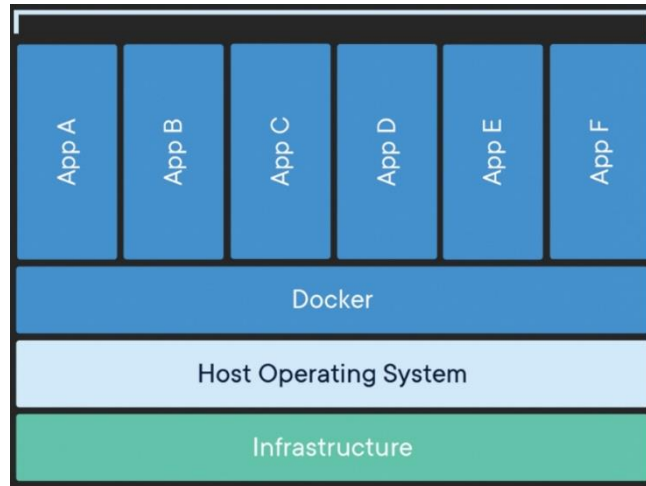
คือต้นแบบของ Container โดยข้างในจะเป็นระบบปฏิบัติการ Linux ที่มีการติดตั้ง Application และ มีการ Configuration เอาไว้ ซึ่งเกิดจากการ build ไฟล์ Docker file ขึ้นมาเป็น image

#### 2) Docker container

Docker container จะถูกสร้างมาจาก Docker Image ที่เป็นต้นแบบหรือแม่พิมพ์ เกิดเป็น container และจะได้ Service หรือ Application ที่สามารถเรียกใช้งานได้ทันที

#### 3) Docker registry

การสร้าง Docker Image แล้วนำไปเก็บรวบรวมไว้บน server (ลักษณะเดียวกับการเก็บ Source Code ไว้บน (Github) โดย Docker registry ณ ปัจจุบันก็มีให้เลือกใช้งานได้อย่างหลากหลายโดยมี Docker Hub เป็น Docker registry หลักในการเรียกใช้(pull) Docker Image และนอกจากนี้ยังมีผู้ให้บริการ docker registry อื่นๆด้วย เช่น Gitlab, Quay.io, Google Cloud เป็นต้น



รูป 2.6 การทำงานของแอปพลิเคชันต่าง ๆ บน Docker Engine

(ที่มา : [docker.com](https://docker.com))

### 2.2.2 Django

Django Framework เป็นชุดของเครื่องมือ Framework สำหรับ การนำไปพัฒนาเว็บไซต์ด้วยภาษาของ Python โดยทุกวันนี้ Framework สำหรับการเขียนเว็บไซต์ด้วยภาษา Python มีค่อนข้างที่จะเยอะ ซึ่ง Django Framework ก็เป็นหนึ่งใน Framework สำหรับการพัฒนาเว็บไซต์ และทำเว็บไซต์ด้วยภาษา Python ด้วยเช่นกัน

คุณสมบัติของ Django Framework

- 1) Object-relational mapper คือ การกำหนด Data Model ในภาษา Python เพื่อใช้ในการทำงานด้านข้อมูล และช่วยสนับสนุน dynamic database-access API
- 2) Automatic admin interface คือ ส่วนในการสร้าง Interface อัตโนมัติสำหรับการ add, edit, delete และ search ด้วย Django Framework
- 3) Elegant URL design คือ การทำให้ URL มีความสั้น กระชับ สวยงาม และสื่อความหมายของหน้านั้น ๆ ได้อย่างชัดเจน
- 4) Template system คือ Django นั้นมีการออกแบบ Template Language เพื่อการเขียนแยกส่วนระหว่าง Design และ Business Logic

- 5) Cache system คือ ส่วนของการบันทึก หรือจัดการข้อมูลที่มีการดาวน์โหลดไปแล้ว เพื่อเพิ่มประสิทธิภาพการทำงานของเว็บไซต์ด้านความเร็ว และด้านอื่น ๆ
- 6) Internationalization คือ Django สนับสนุน Application ที่มีความหลากหลายด้านภาษาในการแสดงผล

### 2.2.3 Scikit-learn

Scikit-learn เป็นโมดูลหนึ่งของภาษา Python เป็นแพ็คเกจที่รวบรวม Library ด้าน การเรียนรู้ของเครื่อง (Machine Learning) เอาไว้ และถูกออกแบบมาให้ทำงานร่วมกับ Library ของภาษา Python อย่าง NumPy และ SciPy ได้ดี

Scikit-learn ยังเป็น Open Source ที่เปิดให้สามารถเข้าไปพัฒนาต่อยอดได้และเป็นแหล่งรวม Library และอัลกอริทึมที่เน้นไปในด้านของ การเรียนรู้ของเครื่อง (Machine Learning) ซึ่งมีส่วนในการทำ แบบจำลองข้อมูล (Data Modeling) อีกหนึ่งปัจจัยที่ทำให้มีผู้ใช้เยอะ เพราะเป็น Interface ระดับสูง ทำให้มือใหม่สามารถเข้าใจภาพรวมและ ขั้นตอนการทำงาน ของการเรียนรู้ของเครื่อง (Machine Learning) ได้เครื่องมือที่ผู้ใช้งานสามารถนำไปใช้ในได้

### 2.2.4 MariaDB

MariaDB คือ เป็น Open Source สำหรับจัดการกับฐานข้อมูล MariaDB เป็นหนึ่งในฐานข้อมูลที่ได้รับความนิยมมากที่สุดในโลก MariaDB ถูกพัฒนาขึ้นโดยนักพัฒนาเดิมของ MySQL เนื่องจากความกังวลที่เกิดขึ้นเมื่อ MySQL ถูกซื้อโดย Oracle Corporation ในปี 2009 ตอนนี้นักพัฒนาและผู้ดูแลของ MariaDB ได้รวมรายเดือนกับฐานรหัส MySQL เพื่อให้แน่ใจว่า MariaDB มีการแก้ไขข้อบกพร่องที่เกี่ยวข้องเพิ่มลงใน MySQL

MariaDB ได้รับการพัฒนาเป็นซอฟต์แวร์โอเพ่นซอร์ส และเป็นฐานข้อมูลเชิงสัมพันธ์แบบ SQL สำหรับการเข้าถึงข้อมูล เวอร์ชันล่าสุดของ MariaDB มีคุณลักษณะ GIS และ JSON ด้วย

MariaDB เปลี่ยนข้อมูลเป็นฐานข้อมูลที่มีโครงสร้างในหลากหลายแอปพลิเคชัน ตั้งแต่ธนาคารไปจนถึงเว็บไซต์ต่างๆ เป็นการปรับปรุงและแทนที่ด้วยการแทนที่ของ MySQL เนื่องจากมีความรวดเร็วและสามารถปรับขนาดได้และมีระบบเวดล๊อคที่อุดมไปด้วยปลั๊กอิน เอนจินและเครื่องมืออื่น ๆ ทำให้สามารถใช้งานได้หลากหลาย

#### **2.2.5 React**

React เป็น JavaScript library ที่ใช้สำหรับสร้าง user interface ที่ให้เราสามารถเขียนโค้ดในการสร้าง UI ที่มีความซับซ้อนแบ่งเป็นส่วนเล็กๆออกจากกันได้ ซึ่งแต่ละส่วนสามารถแยกการทำงานออกจากกันได้อย่างอิสระ และทำให้สามารถนำชิ้นส่วน UI เหล่านั้นไปใช้ซ้ำได้

#### **2.2.6 Node.JS**

Node.js คือสภาพแวดล้อมการทำงานของภาษา JavaScript นอกจากเว็บเบราว์เซอร์ที่ทำงานด้วย V8 engine นั้นหมายความว่าเราสามารถใช้ Node.js ในการพัฒนาแอปพลิเคชันแบบ Command line แอปพลิเคชัน Desktop หรือแม้แต่เว็บเซิร์ฟเวอร์ได้ โดยที่ Node.js จะมี APIs ที่จะสามารถใช้สำหรับทำงานกับระบบปฏิบัติการ เช่น การรับค่าและการแสดงผล การอ่านเขียนไฟล์ และการทำงานกับเน็ตเวิร์ก และยังเป็นโปรแกรมที่สามารถใช้ได้ทั้งบน Windows, Linux และ Mac OS X โดยสามารถเขียนโปรแกรมในภาษา JavaScript และนำไปรันได้ทุกระบบปฏิบัติการที่สนับสนุนโดย Node.js

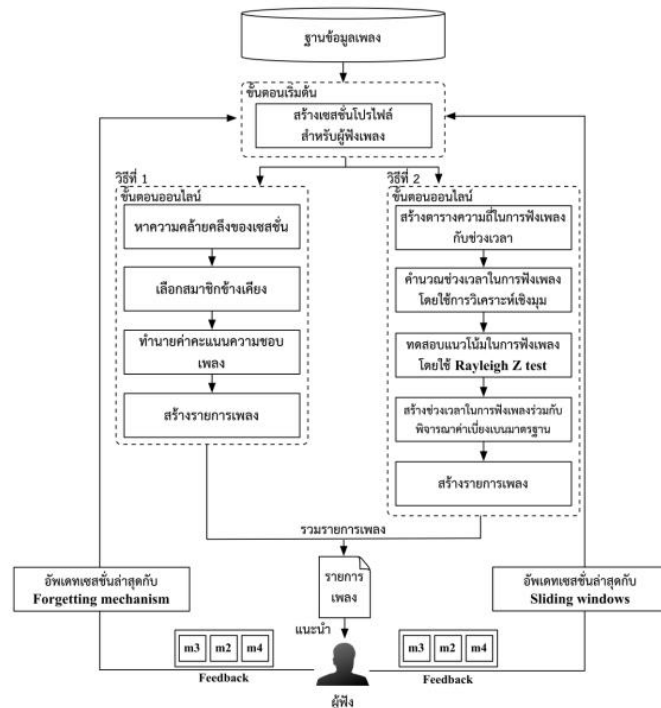


## 2.3 งานวิจัยที่เกี่ยวข้อง

### 2.3.1 การสร้างรายการเพลงโดยใช้การกรองร่วมแบบเซสชันที่เพิ่มขึ้นด้วยกลไกการลื้มและการวิเคราะห์สถิติเชิงมุม

สุเมธ คาราพิสุท นำเสนองานวิจัยเรื่อง การสร้างรายการเพลงโดยใช้การกรองร่วมแบบเซสชันที่เพิ่มขึ้นด้วยกลไกการลื้มและการวิเคราะห์สถิติเชิงมุม โดยใช้ 2 วิธีร่วมกัน 1 การสร้างรายการเพลงจะพิจารณาการฟังเพลงในเซสชันปัจจุบันที่คล้ายกับเซสชันในอดีตของผู้ฟัง 2 สร้างรายการเพลงแนะนำโดยพิจารณาช่วงเวลา เฉพาะในการฟังเพลงซึ่งแตกต่างจากช่วงเวลาอื่นอย่างมีนัยสำคัญทางสถิติในรอบวันของผู้ฟังโดยใช้ การวิเคราะห์สถิติเชิงมุม และวัดประสิทธิภาพโดย ประสิทธิภาพ HitRatio และ Precision จากการทดลองพบว่าการใช้ 2 วิธีแยกกันนั้นได้ผลลัพธ์ที่น้อยกว่านำมาใช้ร่วมกัน 0.18-0.22 % โดยวัตถุประสงค์ในการทำเพื่อวิเคราะห์ปัญหาที่เกิดขึ้นในการสร้างรายการเพลงแนะนำแบบออฟไลน์ พัฒนาขั้นตอนวิธีการสร้างรายการเพลงแนะนำให้มีประสิทธิภาพเพิ่มขึ้นทั้งทางด้านความเร็วและความถูกต้องในการสร้างรายการเพลง

โดยในโครงงานของผู้จัดทำนั้นได้นำส่วนของการออกแบบ Diagram ในงานวิจัยนี้มาใช้งาน โดยใช้วิธีที่ 1 ซึ่งของผู้จัดทำจะเป็น 1 หาความคล้ายคลึงของหมวดหมู่วิชา 2 เลือกสมาชิกข้างเคียง 3 ทำนายค่าผลลัพธ์การเรียนหรือเกรด 4 นำไปสร้างรายการสำหรับขั้นตอนต่อไป



รูป 0.2 ประเภทของ Machine Learning

(ที่มา : ดาราพิศุทธิ์, 2016)

### 2.3.2 การสร้างรายการเพลงโดยใช้การกรองร่วมแบบเซตข้อมูลที่เพิ่มขึ้นด้วยกลไกการลืมและการวิเคราะห์สถิติเชิงมุม

นิภาภรณ์ พันธุ์นาม นำเสนองานวิจัย ระบบแนะนำสินค้าอาหารโดยใช้ระบบแนะนำแบบผสมผสาน ใช้เทคนิค Content based filtering แบบหลักการ Cosine และสร้างแบบจำลองโดยใช้ lib Surprise ซึ่งมีอัลกอริทึม SVD, NMF, Baseline และ KNN และวัดประสิทธิภาพโดย RMSE, MAE จากการทดลองพบว่า 1 เทคนิคการกรองแบบอิงเนื้อหา นำวิธีการ TF-IDF เข้ามาช่วยในการทำ Vectorization ส่วนใหญ่ค่าความเหมือนออกมาค่อนข้างที่จะต่ำเนื่องจากข้อมูลที่น้อยเกินไป 2 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม ผ่าน library Surprise ของ Scikitlearn ซึ่งโมเดลที่มีผลคะแนนโดยรวมดีที่สุดคืออัลกอริทึมของ SVD ซึ่งได้ค่า RMSE 1.2528 และ MAE 0.9376 และ 3 ระบบแนะนำแบบผสมผสาน โดยผลลัพธ์นั้นจะไม่ชัดเจนเนื่องจากวิธีนี้ได้มีการทำนายค่า Rating ซึ่งวิธีการของระบบแนะนำแบบผสมผสานนั้น ได้มีนำเทคนิคการกรองแบบอิงเนื้อหา ที่ไม่ได้มีการทำนายค่าอะไรมารวมในการทำงานของแบบจำลองด้วย ซึ่งถ้าต้องการวัดผลลัพธ์สามารถอ้างอิงจากค่า RMSE, MAE ได้

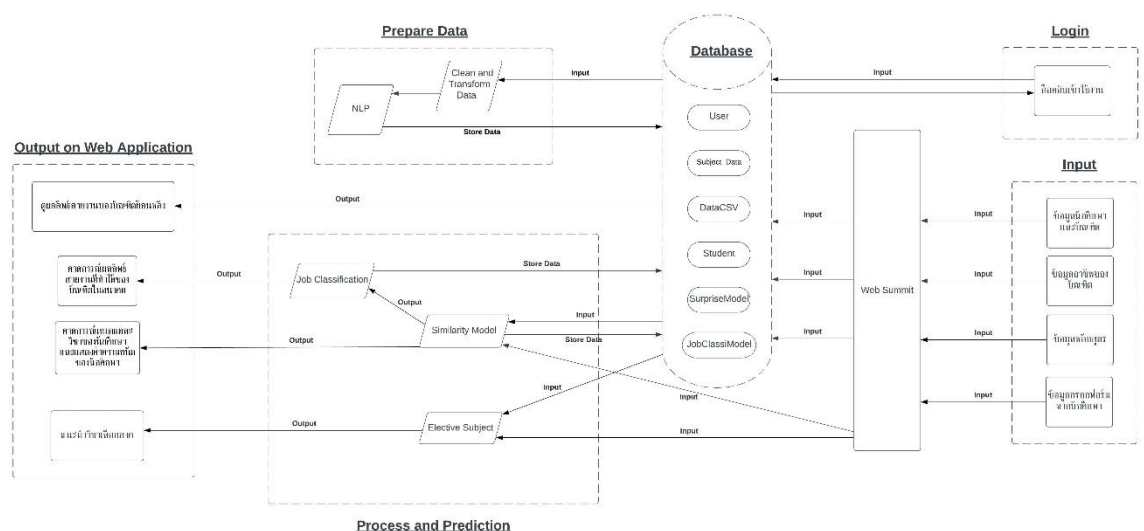
โดยในงานโครงงานของผู้จัดทำนั้นได้นำผลลัพธ์การทดลองของงานวิจัยนี้ที่สรุปได้ว่าผลคะแนนโดยรวมดีที่สุดคืออัลกอริทึม SVD เป็นตัวตัดสินใจในการเลือกใช้อัลกอริทึมนี้และได้นำวิธีการการวัดประสิทธิภาพของแบบจำลองนี้จากงานวิจัยมาปรับใช้ในรูปแบบเดียวกันกับตัวโครงงาน

## บทที่ 3

### การออกแบบ

#### 3.1 โครงสร้างการทำงานของระบบ

โครงสร้างการทำงานของระบบได้อธิบายถึงการเชื่อมต่อระหว่างส่วนต่างๆของระบบ เริ่มตั้งแต่ส่วนของ Input ที่ทำหน้าที่รับข้อมูลเข้ามา แล้วเก็บไว้ในส่วนของ Database จากนั้น ส่วนของ Prepare Data จะนำข้อมูลจากส่วนของ Database เมื่อทำเสร็จแล้ว จะทำการส่งกลับไป อัปเดตยัง Database ส่วนของ Process and Prediction จะนำข้อมูลที่ได้อัปโหลดกลับมาเพื่อ Process ผลลัพธ์ออกมาแสดงผลบน Web Application



รูป 3.1 โครงสร้างการทำงานของระบบ

#### 3.2 การทำงานของระบบ

จากรูป 1 โครงสร้างการทำงานของระบบนั้นประกอบไปด้วยองค์ประกอบทั้งหมด 6 ส่วน ซึ่งมีรายละเอียด ดังนี้

## 1. Login

เป็นส่วนสำหรับไว้ให้กรรมการหลักสูตรได้ทำการล็อกอินเข้าสู่ระบบเพื่อให้กรรมการหลักสูตรป้อนข้อมูลเกรดของนักศึกษาปัจจุบัน เกรดและอาชีพของบัณฑิต และข้อมูลของหลักสูตร

## 2. Input

เป็นส่วนที่ทำการรับข้อมูลของนักศึกษาและบัณฑิต ข้อมูลของหลักสูตร แล้วเก็บเข้ายังส่วนของ Database และข้อมูลการกรอกฟอร์มของนักศึกษาจะส่งข้อมูลไปยังส่วนของ Process and Prediction โดยตรง

## 3. Database

ทำหน้าที่ในการจัดเก็บและบันทึกข้อมูล โดยจะประกอบไปด้วย Table User, Subject\_Data, DataCSV, Graduate, Student

## 4. Prepare Data

ทำหน้าที่เตรียมพร้อมข้อมูลเพื่อให้พร้อมต่อการนำไปใช้ในส่วนของ Prediction and Prediction โดยแบ่งออกเป็น 2 ส่วนย่อย ดังนี้

### 1. Clean and Transform Data

เป็นการเรียกข้อมูลจากใน Database มาทำให้ข้อมูลมีความถูกต้อง ครบถ้วน สมบูรณ์ ไม่มีค่าที่ผิดปกติ เพื่อเตรียมพร้อมสำหรับการประมวลผลข้อมูล โดยเลือกเฉพาะข้อมูลที่เกี่ยวข้องกับวัตถุประสงค์ของการวิเคราะห์ และประมวลผล พร้อมทั้งตัดข้อมูลส่วนที่ไม่ได้นำมาใช้ออก

### 2. NLP หรือ Natural Language Processing

นำข้อมูลรหัสวิชามาเข้ากระบวนการ NLP เพื่อทำการหาค่า Similarity ของกลุ่ม วิชาที่สามารถอยู่ในกลุ่มเดียวกันได้ โดยใช้บทคัดย่อของแต่ละวิชา เพื่อลดปัญหา การเปลี่ยนรหัสวิชาระหว่างหลักสูตร เมื่อทำเสร็จกระบวนการแล้ว จะนำข้อมูลที่ ได้ กลับไปอัปเดตที่ Database

## 5. Process and Prediction

เป็นส่วนการประมวลผลหลักของระบบ ประกอบไปด้วย Process 3 ส่วนดังนี้

### 1. Similarity Model

มีหน้าที่นำข้อมูลผลการเรียนของนักศึกษามาทำให้สมบูรณ์ ด้วยกรณีของนักศึกษา แต่ละคนมีเกรดแต่ละวิชาที่ไม่เหมือนกัน บางคนมีเกรดของวิชานี้ แต่อีกคนไม่มี เกรดของวิชานี้ จึงต้องทำให้ข้อมูลของนักศึกษามีเท่ากันเพื่อนำไปทำการ Prediction ในขั้นต่อไป โดยใช้กระบวนการหลักคือการทำ Recommender Systems และใช้ Library Surprise ของ Scikit ในภาษา Python เป็นตัวช่วย

### 2. Job Classification

มีหน้าที่ทำนายและสถิติด้านความสามารถทางวิชาชีพของนักศึกษาในอนาคต

### 3. Elective Subject

มีหน้าที่ประมวลผลจัดกลุ่มของวิชาเลือกภาค เพื่อนำไปแนะนำให้แก่นักศึกษาที่มีความสนใจเฉพาะจุดได้

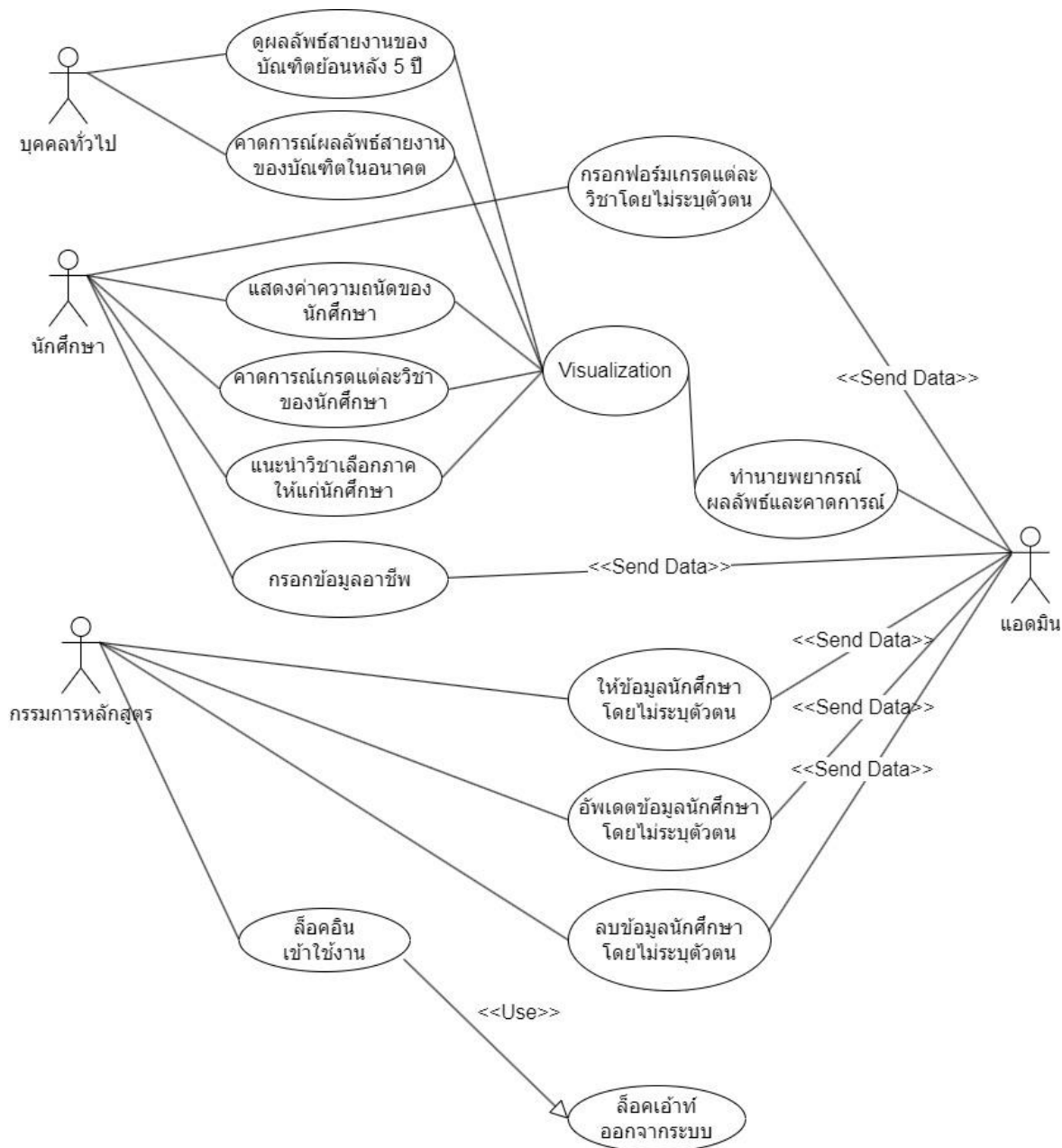
## 6. Web Application

เป็นส่วนที่ทำหน้าที่ในการแสดงผลข้อมูลและเป็น interface สำหรับผู้ใช้งาน

### 3.3 Use Case Diagram

การใช้งานระบบจะแบ่งผู้ใช้ออกเป็น 4 ประเภท ได้แก่ บุคคลทั่วไปหรือนักศึกษา กรรมการหลักสูตร และ แอดมิน

1. บุคคลทั่วไปหรือนักศึกษา มีสิทธิ์เข้าถึงการดูข้อมูลการทำนายผล
2. กรรมการหลักสูตร มีสิทธิ์เข้าถึงในการส่งไฟล์ข้อมูลของนักศึกษา
3. แอดมิน มีสิทธิ์เข้าถึงการทำงานทั้งหมดของระบบวิเคราะห์ และ พยากรณ์



รูป 3.3 Use Case Diagram

ตาราง 3.1 Use Case คู่มือผลลัพธ์รายงานของบัณฑิตย้อนหลัง

<b>Use Case:</b> คู่มือผลลัพธ์รายงานของบัณฑิตย้อนหลัง
<b>Use Case ID:</b> UC-01
<b>Actor:</b> บุคคลทั่วไปหรือนักศึกษา
<b>Description:</b> สามารถดูผลลัพธ์รายงานของบัณฑิตย้อนหลังได้
<b>Precondition:</b> เข้า web application หน้าผลลัพธ์รายงานของบัณฑิตย้อนหลัง
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าคู่มือผลลัพธ์รายงานของบัณฑิตย้อนหลัง</li> <li>2. เลือกปีที่ต้องการดู</li> <li>3. เลือกหลักสูตรที่ต้องการดู</li> </ol>
<b>Postcondition:</b> หน้าเว็บแสดงผลผลลัพธ์รายงานของบัณฑิตย้อนหลัง

ตาราง 3.2 Use Case คาดการณ์ผลลัพธ์รายงานที่ทำได้ของบัณฑิตในอนาคต

<b>Use Case:</b> คาดการณ์ผลลัพธ์รายงานที่ทำได้ของบัณฑิตในอนาคต
<b>Use Case ID:</b> UC-02
<b>Actor:</b> บุคคลทั่วไปหรือนักศึกษา
<b>Description:</b> สามารถดูผลลัพธ์รายงานที่ทำได้ของบัณฑิตในอนาคตได้
<b>Precondition:</b> เข้า web application หน้าผลลัพธ์รายงานที่ทำได้ของบัณฑิตในอนาคต
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าคู่มือผลลัพธ์รายงานที่ทำได้ของบัณฑิตในอนาคต</li> <li>2. เลือกปีที่ต้องการดู</li> <li>3. เลือกหลักสูตรที่ต้องการดู</li> </ol>
<b>Postcondition:</b> หน้าเว็บแสดงผลผลลัพธ์รายงานที่ทำได้ของบัณฑิตในอนาคต



ตาราง 3.3 Use Case กรอกแบบฟอร์มเกรดสำหรับการวิเคราะห์ข้อมูล

<b>Use Case:</b> กรอกแบบฟอร์มเกรดสำหรับการวิเคราะห์ข้อมูล
<b>Use Case ID:</b> UC-03
<b>Actor:</b> บุคคลทั่วไปหรือนักศึกษา
<b>Description:</b> สามารถกรอกแบบฟอร์มการวิเคราะห์ข้อมูลได้
<b>Precondition:</b> เข้า web application หน้าสำหรับการวิเคราะห์ข้อมูลของนักศึกษา
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าสำหรับการวิเคราะห์ข้อมูลของนักศึกษา</li> <li>2. เลือกปีการศึกษา</li> <li>3. เลือกหลักสูตร</li> <li>4. โหลดแบบฟอร์มสำหรับการวิเคราะห์ข้อมูล</li> <li>5. อัปโหลดแบบฟอร์มสำหรับการวิเคราะห์ข้อมูล</li> </ol>
<b>Postcondition:</b> Redirect ไปหน้าแสดงผลวิเคราะห์ข้อมูลนักศึกษา

ตาราง 3.4 Use Case แสดงค่าความถนัดของนักศึกษา

<b>Use Case:</b> แสดงค่าความถนัดของนักศึกษา
<b>Use Case ID:</b> UC-04
<b>Actor:</b> บุคคลทั่วไปหรือนักศึกษา
<b>Description:</b> สามารถดูผลแสดงค่าความถนัดของนักศึกษาได้
<b>Precondition:</b> เข้า web application หน้าแสดงผลวิเคราะห์ข้อมูลนักศึกษา
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าแสดงผลสำหรับการวิเคราะห์ข้อมูลของนักศึกษา</li> </ol>
<b>Postcondition:</b> หน้าเว็บแสดงผลผลลัพธ์การวิเคราะห์ข้อมูลค่าความถนัดของนักศึกษา

ตาราง 3.5 Use Case คาดการณ์เกรดแต่ละวิชาของนักศึกษา

<b>Use Case:</b> คาดการณ์เกรดแต่ละวิชาของนักศึกษา
<b>Use Case ID:</b> UC-05
<b>Actor:</b> บุคคลทั่วไปหรือนักศึกษา
<b>Description:</b> สามารถดูผลแสดงค่าคาดการณ์เกรดแต่ละวิชาของนักศึกษาได้
<b>Precondition:</b> เข้า web application หน้าแสดงผลวิเคราะห์ข้อมูลนักศึกษา
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าแสดงผลสำหรับการวิเคราะห์ข้อมูลของนักศึกษา</li> </ol>
<b>Postcondition:</b> หน้าเว็บแสดงผลผลลัพธ์การวิเคราะห์ข้อมูลค่าคาดการณ์เกรดแต่ละวิชาของนักศึกษา

ตาราง 3.6 Use Case แนะนำวิชาเลือกภาคให้แก่นักศึกษา

<b>Use Case:</b> แนะนำวิชาเลือกภาคให้แก่นักศึกษา
<b>Use Case ID:</b> UC-06
<b>Actor:</b> บุคคลทั่วไปหรือนักศึกษา
<b>Description:</b> สามารถดูผลแนะนำวิชาเลือกให้ภาคให้แก่นักศึกษาได้
<b>Precondition:</b> เข้า web application หน้าแนะนำวิชาเลือกภาค
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าแนะนำวิชาเลือกภาค</li> <li>2. เลือกสิ่งที่สนใจภายในตัวเลือกที่มีให้</li> </ol>
<b>Postcondition:</b> หน้าเว็บแนะนำวิชาเลือกภาค

ตาราง 3.7 Use Case ให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร

<b>Use Case:</b> ให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Use Case ID:</b> UC-07
<b>Actor:</b> กรรมการหลักสูตร
<b>Description:</b> ให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตรได้
<b>Precondition:</b> เข้า web application หน้าให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร</li> <li>2. อัปโหลดไฟล์ CSV</li> <li>3. กดปุ่มอัปโหลด</li> </ol>
<b>Postcondition:</b> อัปโหลดข้อมูลสำเร็จ

ตาราง 3.8 Use Case อัปเดตข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร

<b>Use Case:</b> อัปเดตข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Use Case ID:</b> UC-08
<b>Actor:</b> กรรมการหลักสูตร
<b>Description:</b> อัปเดตนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตรได้
<b>Precondition:</b> เข้า web application หน้าให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร</li> <li>2. อัปเดตไฟล์ CSV</li> <li>3. กดปุ่มอัปโหลด</li> </ol>
<b>Postcondition:</b> อัปเดตข้อมูลสำเร็จ

ตาราง 3.9 Use Case ลบข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร

<b>Use Case:</b> อัปเดตข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Use Case ID:</b> UC-09
<b>Actor:</b> กรรมการหลักสูตร
<b>Description:</b> ลบนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตรได้
<b>Precondition:</b> เข้า web application หน้าให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. เลือกหน้าให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร</li> <li>2. ลบไฟล์ CSV</li> <li>3. กดปุ่มยืนยัน</li> </ol>
<b>Postcondition:</b> ลบข้อมูลสำเร็จ

ตาราง 3.10 Use Case ล็อคอินใช้งาน

<b>Use Case:</b> ให้ข้อมูลนักศึกษาโดยไม่ระบุตัวตนและข้อมูลหลักสูตร
<b>Use Case ID:</b> UC-10
<b>Actor:</b> กรรมการหลักสูตร
<b>Description:</b> กรรมการหลักสูตรสามารถล็อคอินใช้งานได้
<b>Precondition:</b> เข้า web application หน้า login
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. กรอกข้อมูล username และ password</li> <li>2. กดปุ่ม Login</li> </ol>
<b>Postcondition:</b> Redirect ไปหน้าหลัก

ตาราง 3.11 Use Case ทำนายพยากรณ์ผลลัพธ์และคาดการณ์

<b>Use Case:</b> ทำนายพยากรณ์ผลลัพธ์และคาดการณ์
<b>Use Case ID:</b> UC-11
<b>Actor:</b> แอดมิน
<b>Description:</b> แอดมินสามารถทำนายพยากรณ์ผลลัพธ์และคาดการณ์ได้
<b>Precondition:</b> เป็นแอดมิน
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. ทำการทำนายพยากรณ์และคาดการณ์</li> <li>2. ส่งออกผลลัพธ์ไปยังหน้าเว็บ</li> </ol>
<b>Postcondition:</b> ทำนายพยากรณ์ผลลัพธ์และคาดการณ์ได้สำเร็จ

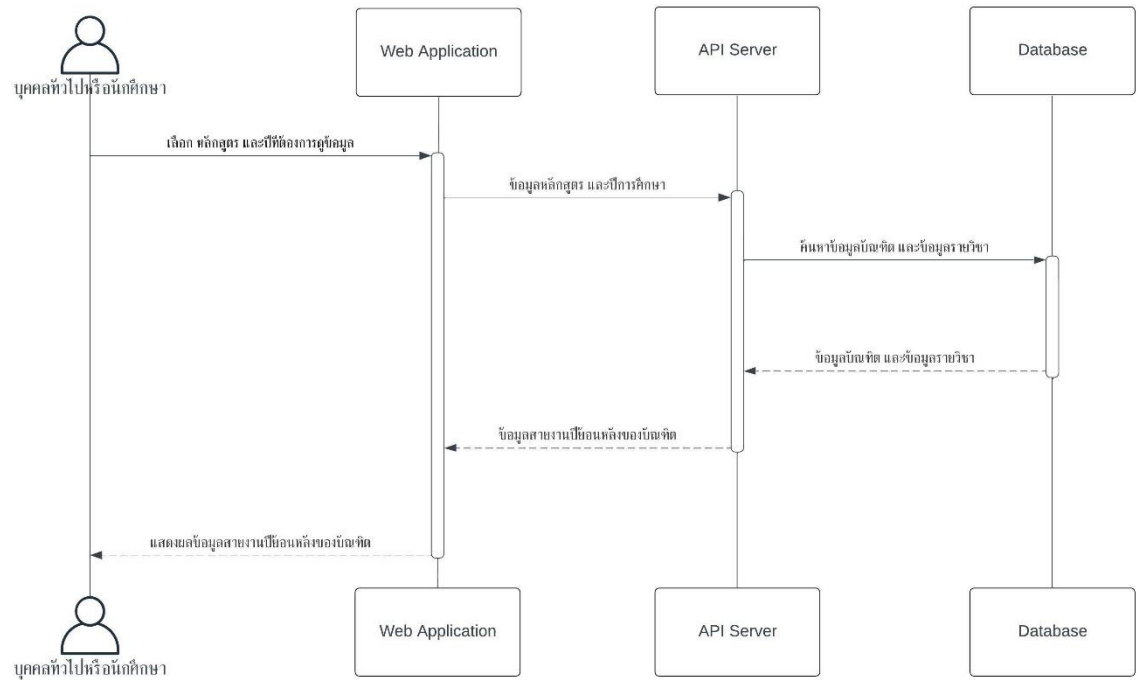
ตาราง 3.12 Use Case จัดเตรียมข้อมูล

<b>Use Case:</b> จัดเตรียมข้อมูล
<b>Use Case ID:</b> UC-12
<b>Actor:</b> แอดมิน
<b>Description:</b> จัดเตรียมข้อมูลสำหรับการทำนายและคาดการณ์ได้
<b>Precondition:</b> เป็นแอดมิน
<b>Flow of Events:</b> <ol style="list-style-type: none"> <li>1. นำเข้าข้อมูล</li> <li>2. จัดรูปแบบข้อมูล</li> <li>3. เก็บเข้าฐานข้อมูล</li> </ol>
<b>Postcondition:</b> จัดเตรียมข้อมูลสำหรับการทำนายและคาดการณ์ได้สำเร็จ

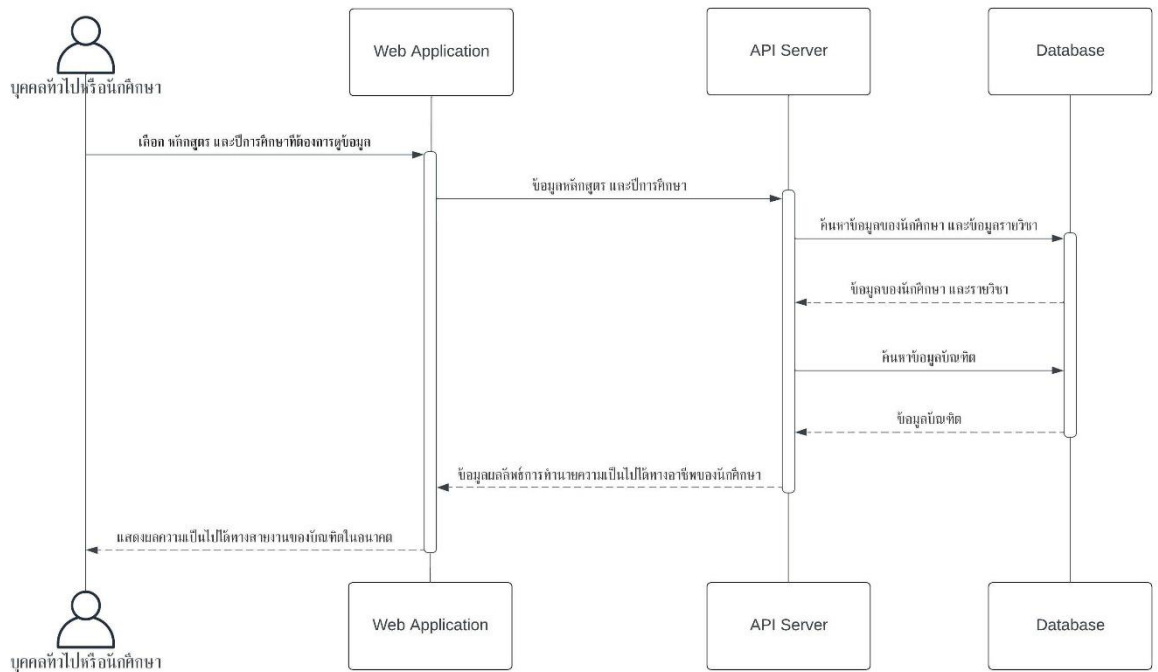
**ตาราง 3.13 Use Case กรอกข้อมูลอาชีพ**

<b>Use Case:</b> กรอกข้อมูลอาชีพ
<b>Use Case ID:</b> UC-13
<b>Actor:</b> นักศึกษา
<b>Description:</b> กรอกข้อมูลอาชีพแรกของการทำงานได้
<b>Precondition:</b> เข้า web application หน้ากรอกข้อมูลอาชีพ
<b>Flow of Events:</b> <ol style="list-style-type: none"><li>1. เลือกหน้ากรอกข้อมูลอาชีพ</li><li>2. ใส่รหัสนักศึกษา</li><li>3. ใส่ตำแหน่งอาชีพ</li><li>4. กดยืนยัน</li></ol>
<b>Postcondition:</b> กรอกข้อมูลอาชีพสำเร็จ

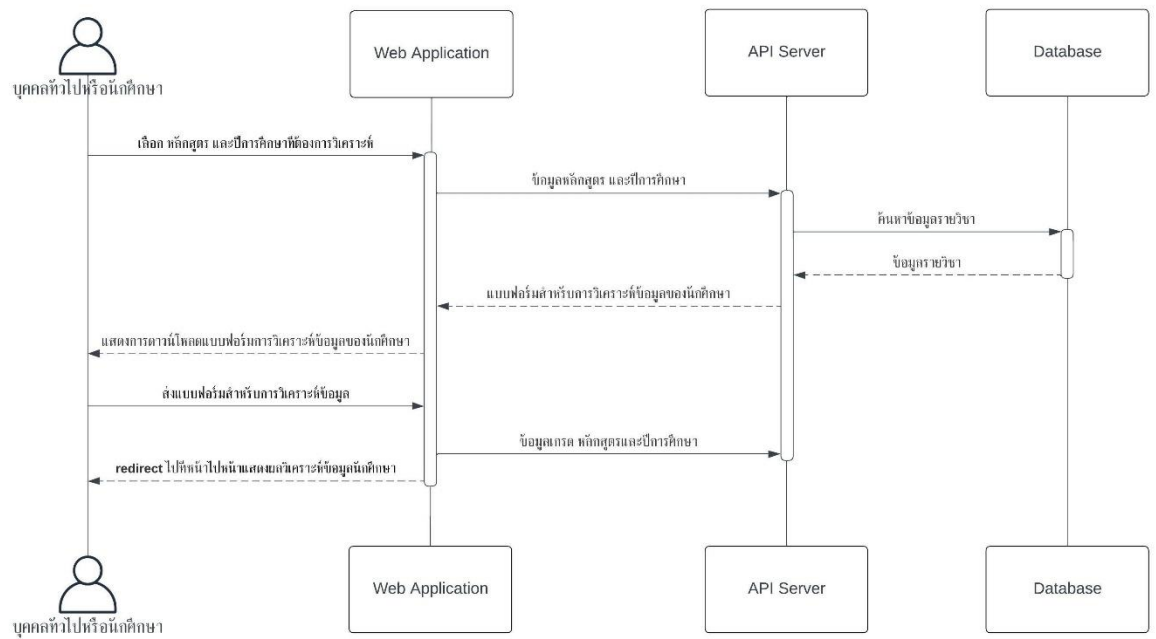
### 3.4 Sequence Diagram



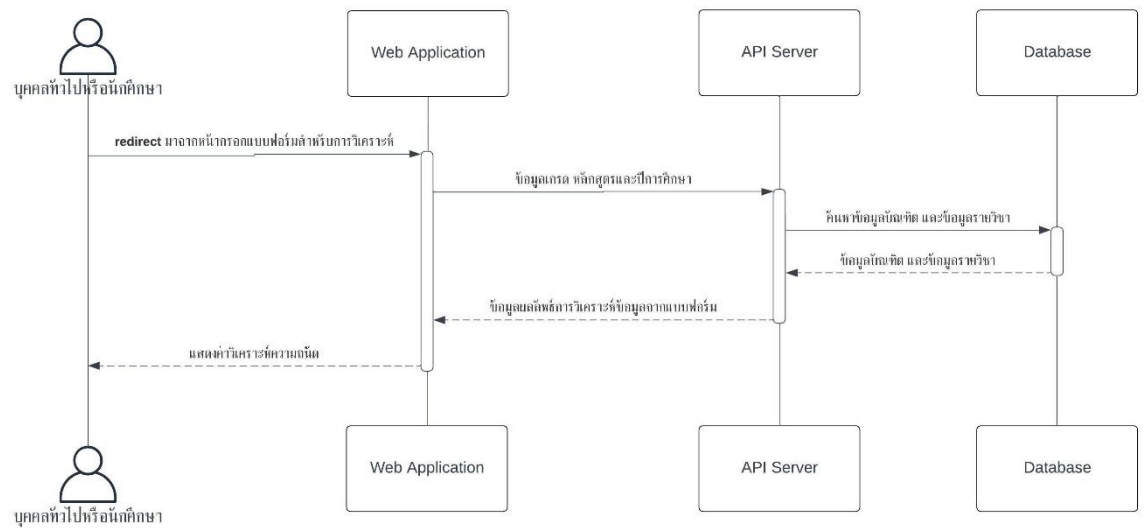
รูป 3.4 Sequence Diagram สำหรับ UC-01



รูป 3.5 Sequence Diagram สำหรับ UC-02

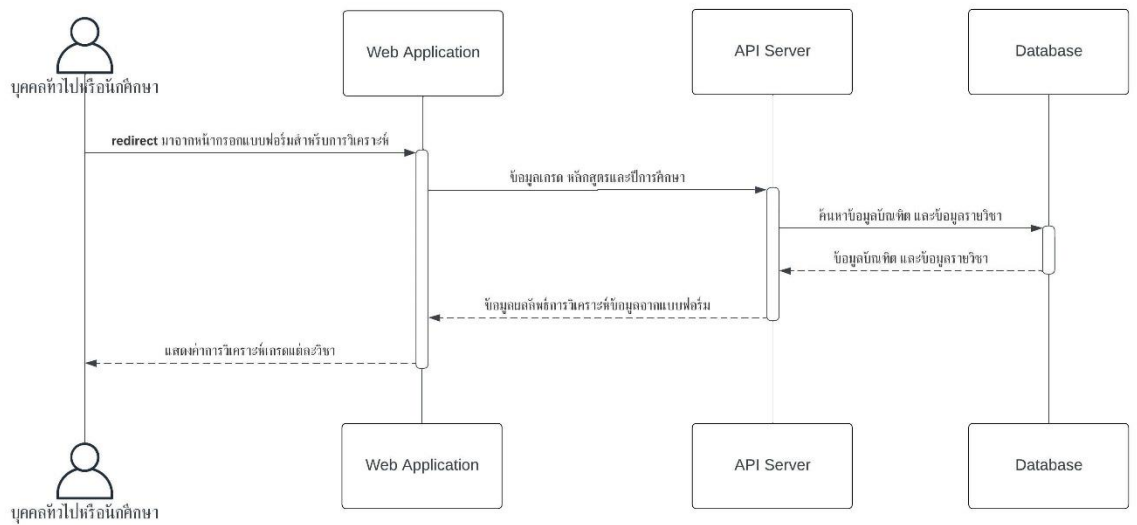


รูป 3.6 Sequence Diagram สำหรับ UC-03

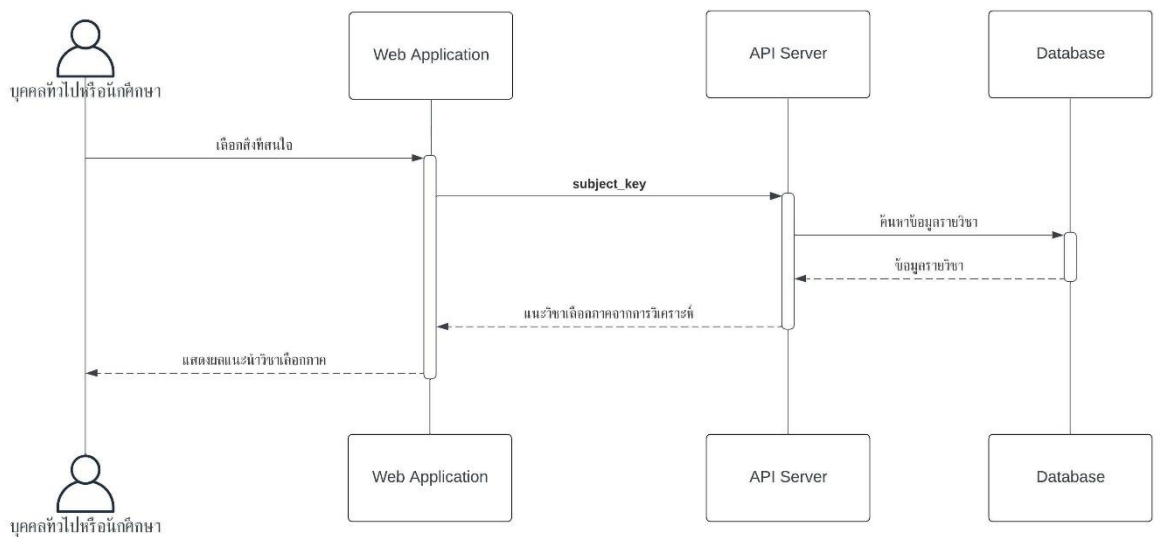


รูป 3.7 Sequence Diagram สำหรับ UC-04

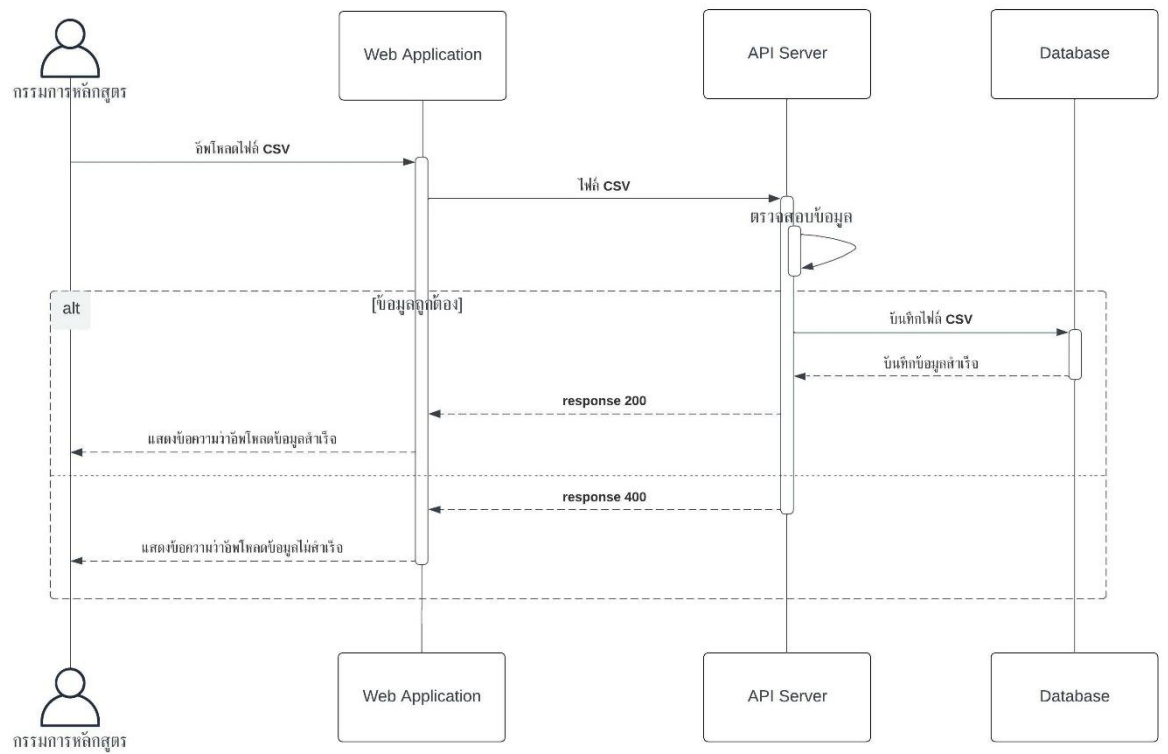




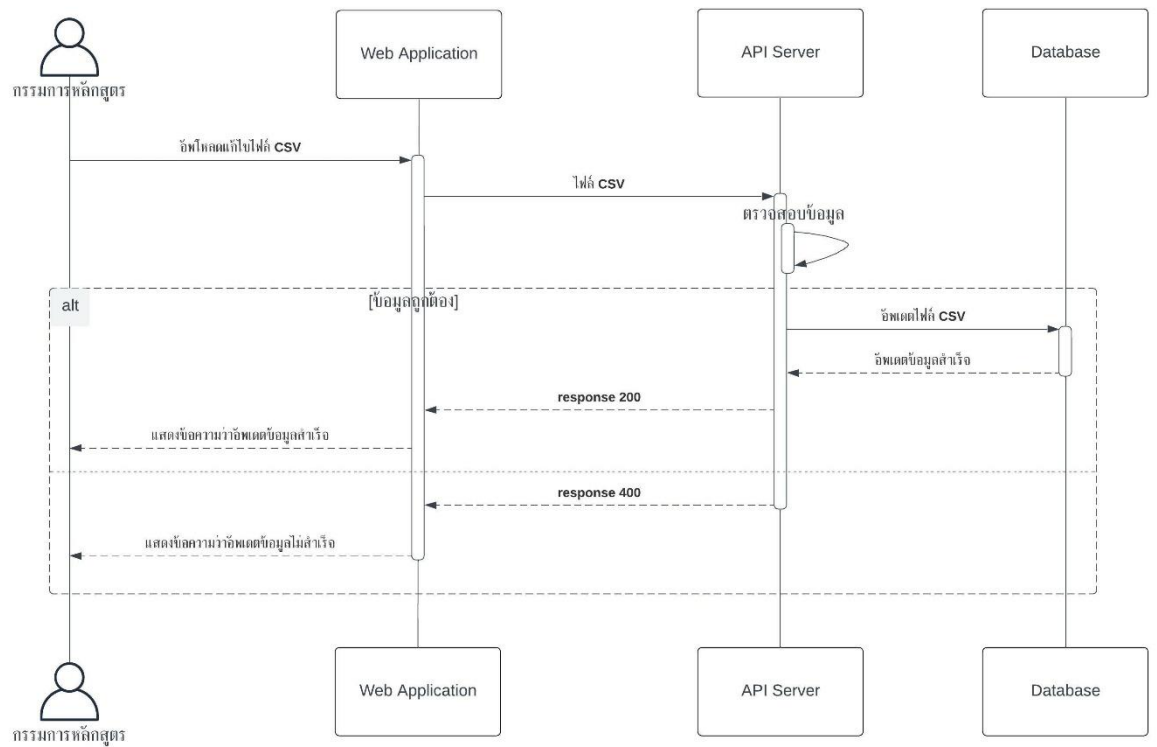
รูป 3.8 Sequence Diagram สำหรับ UC-05



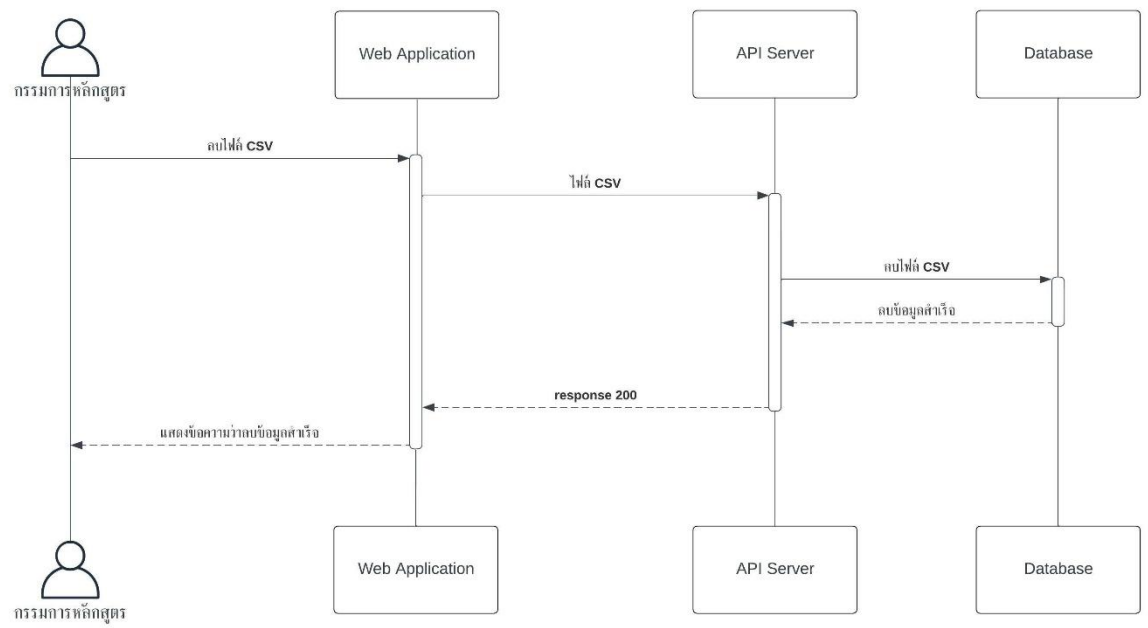
รูป 3.9 Sequence Diagram สำหรับ UC-06



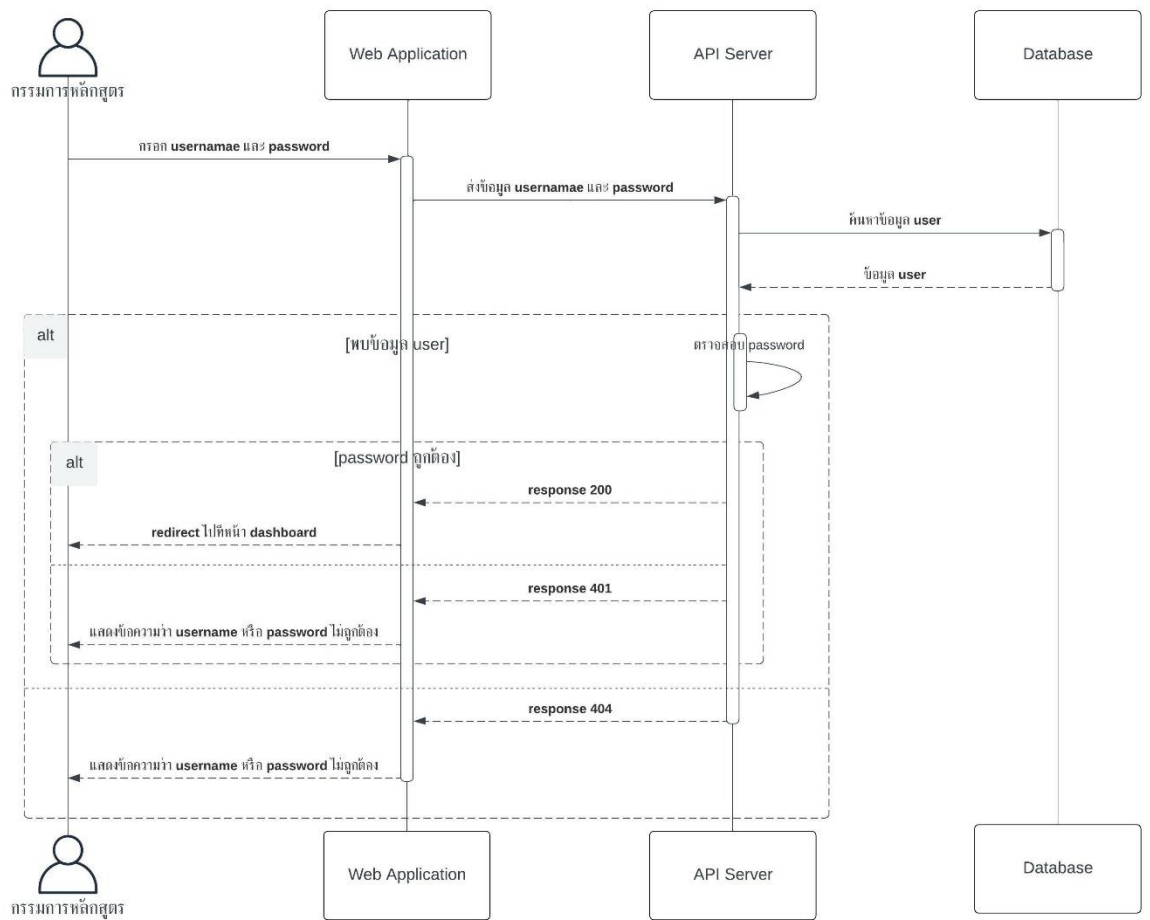
รูป 3.10 Sequence Diagram สำหรับ UC-07



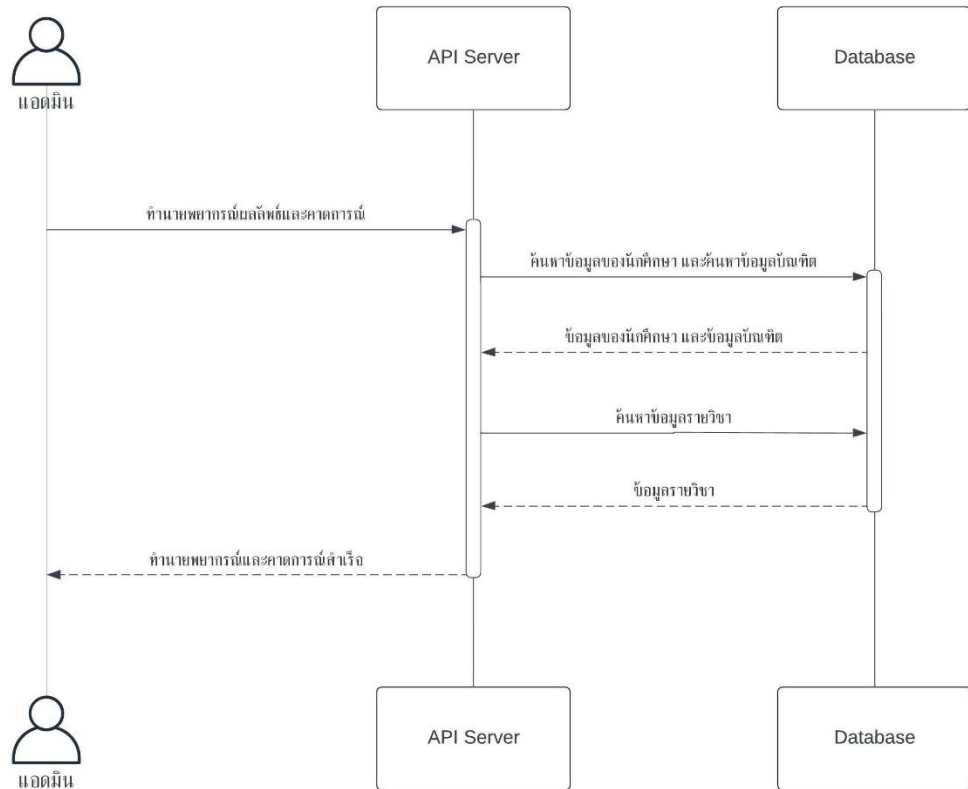
รูป 3.11 Sequence Diagram สำหรับ UC-08



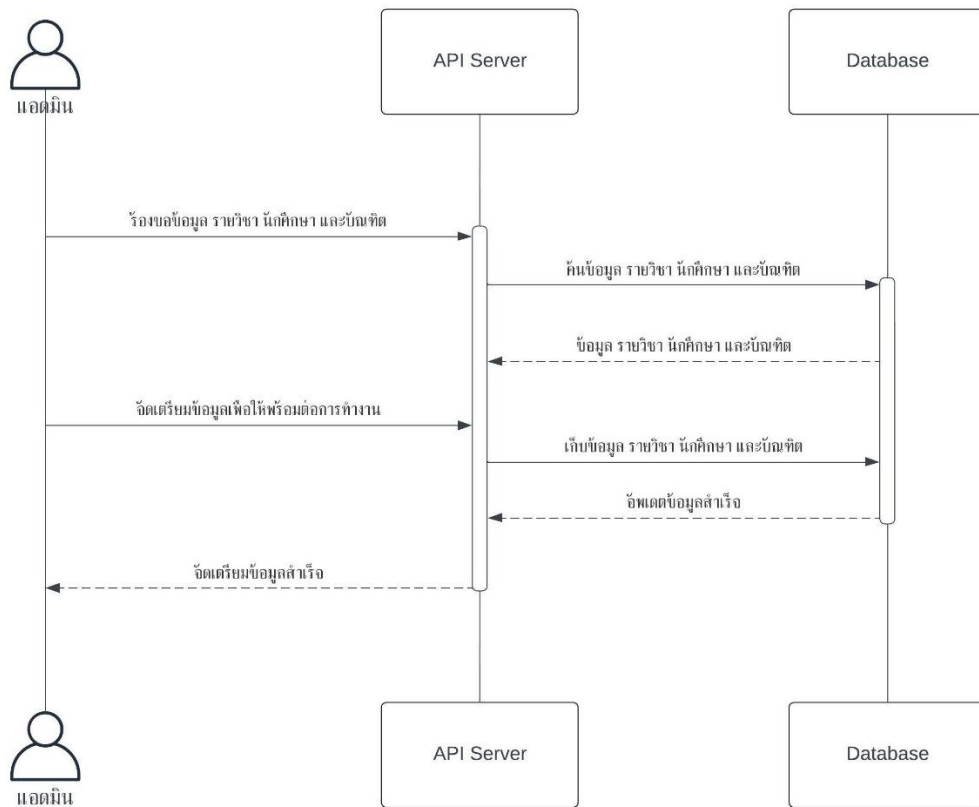
รูป 3.12 Sequence Diagram สำหรับ UC-09



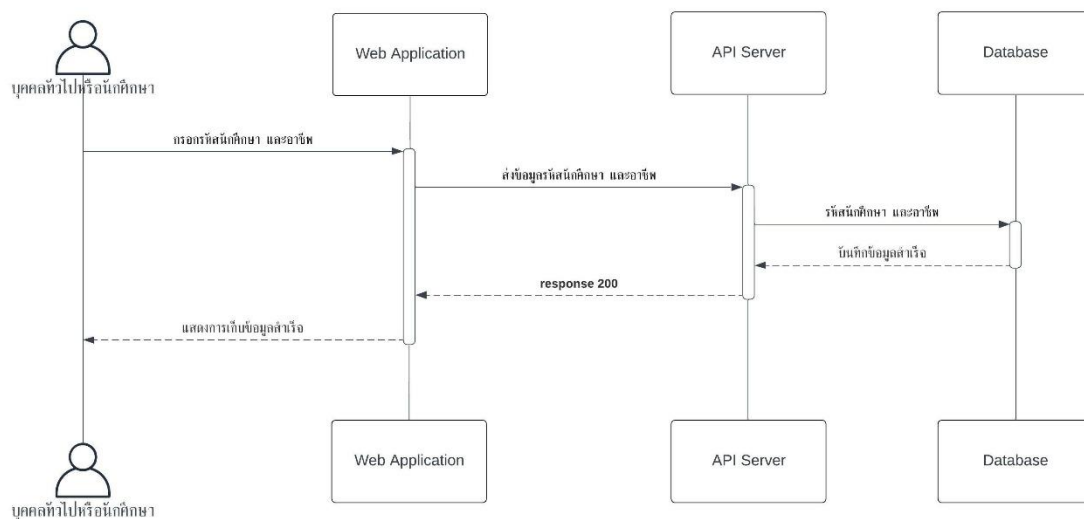
รูป 3.13 Sequence Diagram สำหรับ UC-10



รูป 3.14 Sequence Diagram สำหรับ UC-11



รูป 3.15 Sequence Diagram สำหรับ UC-12



รูป 3.16 Sequence Diagram สำหรับ UC-13

### 3.5 การออกแบบฐานข้อมูล

ระบบฐานข้อมูลของโครงการนี้ผู้จัดทำได้เลือกใช้ MariaDB ซึ่งเป็นฐานข้อมูลแบบ SQL ที่ถูกพัฒนาขึ้นมาเพื่อสำหรับเก็บข้อมูล โดยฐานข้อมูลของระบบประกอบไปด้วยทั้งหมด 5 ตาราง ดังรูป

#### 3.5.1 ตาราง User

เป็นตารางที่ใช้เก็บข้อมูลผู้ใช้งานในระบบ โดยจะมี Attribute ดังนี้

1. user\_id : UUID สำหรับเก็บ id ของผู้ใช้
2. username : VARCHAR สำหรับเก็บ username ของผู้ใช้
3. password : VARCHAR สำหรับเก็บ password ของผู้ใช้

#### 3.5.2 ตาราง Subject\_Data

เป็นตารางที่ใช้เก็บข้อมูลหลักสูตรวิชา โดยจะมี Attribute ดังนี้

1. subject\_id : UUID สำหรับเก็บ id ของรหัสวิชา
2. subject\_name\_thai : VARCHAR สำหรับเก็บชื่อรายวิชาภาษาไทย
3. subject\_name\_eng : VARCHAR สำหรับเก็บชื่อรายวิชาภาษาอังกฤษ
4. abstract : VARCHAR สำหรับเก็บบทคัดย่อรายวิชา

5. subject\_key : VARCHAR สำหรับเก็บ Keyword ของรายวิชาเลือกภาค
6. subject\_class : VARCHAR สำหรับเก็บหมวดหมู่ของแต่ละวิชา
7. year : VARCHAR สำหรับเก็บปีของเล่มหลักสูตรวิชา

### 3.5.3 ตาราง Student

เป็นตารางที่ใช้เก็บข้อมูลนักศึกษาโดยจะมี Attribute ดังนี้

1. student\_id : UUID สำหรับเก็บ id encrypt ของนักศึกษา
2. subject\_id : VARCHAR สำหรับเก็บ id ของรหัสวิชา
3. grade : VARCHAR สำหรับเก็บเกรดแต่ละรายวิชา
4. semester : VARCHAR สำหรับเก็บเทอมที่ลงเรียนรายวิชาของนักศึกษา
5. year : VARCHAR สำหรับเก็บปีที่ลงเรียนรายวิชาของนักศึกษา
6. curriculum : VARCHAR สำหรับเก็บรายชื่อหลักสูตรที่เรียนของนักศึกษา
7. career : VARCHAR สำหรับเก็บรายชื่อตำแหน่งอาชีพแรกของนักศึกษา
8. status : VARCHAR สำหรับเก็บสถานะการเป็นนักศึกษา

### 3.5.4 ตาราง DataCSV

เป็นตารางที่ใช้เก็บข้อมูลไฟล์ CSV โดยจะมี Attribute ดังนี้

1. data\_id : UUID สำหรับเก็บ id ของไฟล์ CSV
2. name : VARCHAR สำหรับเก็บชื่อของไฟล์ CSV
3. upload\_date : TIMESTAMP สำหรับเก็บเวลาและวันที่อัปโหลดไฟล์ CSV
4. update\_date : TIMESTAMP สำหรับเก็บเวลาและวันที่อัปเดตไฟล์ CSV
5. del\_flag : VARCHAR สำหรับเก็บสถานะไฟล์ CSV (0 = ไฟล์ยังมีอยู่, 1 = ไฟล์ถูกลบ)

### 3.5.5 ตาราง SurpriseModel

เป็นตารางที่ใช้เก็บ Surprise Model โดยจะมี Attribute ดังนี้

1. model\_id : UUID สำหรับเก็บ id ของไฟล์ Model
2. name : VARCHAR สำหรับเก็บชื่อของ Model
3. curriculum : VARCHAR สำหรับเก็บรายชื่อหลักสูตรที่เรียนของนักศึกษา
4. type : VARCHAR สำหรับเก็บประเภทของ Model
5. rmse : VARCHAR สำหรับเก็บค่าความแม่นยำของ Model
6. model : **PICKLED** สำหรับเก็บตัว Model

### 3.5.6 ตาราง JobClassiModel

เป็นตารางที่ใช้เก็บ Job Classification Model โดยจะมี Attribute ดังนี้

1. `model_id` : UUID สำหรับเก็บ id ของไฟล์ Model
2. `name` : VARCHAR สำหรับเก็บชื่อของ Model
3. `curriculum` : VARCHAR สำหรับเก็บรายชื่อหลักสูตรที่เรียนของนักศึกษา
4. `type` : VARCHAR สำหรับเก็บประเภทของ Model
5. `rmse` : VARCHAR สำหรับเก็บค่าความแม่นยำของ Model
6. `model` : PICKLED สำหรับเก็บตัว Model