



**ข้อเสนอโครงการวิศวกรรมคอมพิวเตอร์**  
**วิชา 01076014 การเตรียมโครงการวิศวกรรมคอมพิวเตอร์**  
**ภาคเรียนที่ 2 ปีการศึกษา 2564**

1. ชื่อหัวข้อโครงการ (ไทย) .....ระบบแนะนำสำหรับการบริหารหลักสูตรวิศวกรรมคอมพิวเตอร์.....
2. ชื่อหัวข้อโครงการ (อังกฤษ) .....Recommendation System for CE Curriculum administrators.....
3. Keyword 3 คำ .....  
 4. ประเภทโครงการ (✓)  

☐ 1. HW+SW

☒ 2. SW\_Dev

☐ 3. Research
5. รายชื่อผู้ทำโครงการ
  - 5.1. นางสาว .....ณิกานต์ สุขุมจิตพิทย.....รหัส .....62010299.....
  - 5.2. นาย.....นนทพันธุ์ รุจิรกาล.....รหัส .....62010453.....
  - 5.3. นาย.....นรวิชญ์ อยู่บัว.....รหัส .....62010465.....
6. อาจารย์ที่ปรึกษา
  - 6.1. อาจารย์ที่ปรึกษาหลัก .....ผศ. ดร. จันญชัย ตรีภาค.....
 อาจารย์ที่ปรึกษาร่วม .....

## 1. ที่มาและความสำคัญของปัญหา (Motivation)

Data Analytics เป็นการวิเคราะห์ข้อมูลที่มีอยู่ตั้งแต่อดีตจนถึงปัจจุบัน ในกรณีที่ข้อมูลเพียงพอและเหมาะสมจะสามารถนำมาคาดการณ์แนวโน้ม ทำนายอนาคตที่เป็นประโยชน์ พยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้นโดยใช้ข้อมูลในอดีตกับแบบจำลองทางสถิติ รวมถึงการให้คำแนะนำทางเลือกต่าง ๆ และผลของแต่ละทางเลือก

จากปัญหาที่ทางผู้จัดทำเล็งเห็นความสำคัญคือการนำข้อมูลผลการเรียนของนักศึกษาในอดีตมาใช้ประโยชน์ในการบริหารหลักสูตร และ นำมาวิเคราะห์ผลเพื่อช่วยในการวางแผนการเรียนของนักศึกษา ซึ่งการวางแผนในการเรียนของหลักสูตรจะสามารถช่วยอาจารย์และบุคลากรที่เกี่ยวข้องกับการศึกษาในด้านการบริหารหลักสูตร เพื่อวางแผนการเพิ่มหรือลดจำนวนผู้เรียนในรายวิชาต่าง ๆ ซึ่งส่งผลต่อการผลิตบัณฑิตด้านต่าง ๆ ได้

ดังนั้นผู้จัดทำจึงได้เห็นถึงความสำคัญการประเมินสถานะของหลักสูตร ของระบบแนะนำการวางแผน การคาดการณ์ จากการใช้ความรู้ทางด้าน Data Analytics, Prediction และระบบ Data Visualization โดยใช้ข้อมูลผลการเรียนของนักศึกษาในอดีต เพื่อพัฒนาระบบช่วยเหลือ และตอบโจทย์ให้แก่นักศึกษาและบุคลากรทางการศึกษาหรือบุคคลที่เกี่ยวข้องได้

## 2. วัตถุประสงค์ (Objectives)

1. เพื่อนำข้อมูลของผลการเรียนของนักศึกษาในอดีตและข้อมูลจากแบบสำรวจการปฏิบัติงานของบัณฑิตมาใช้ในการวางแผนการเรียนหรือประเมินอาชีพในอนาคตของนักศึกษาได้
2. ประมวลผลข้อมูลผลการเรียนของนักศึกษาในอดีต และข้อมูลจากแบบสำรวจการปฏิบัติงานของบัณฑิต และทำแผนภาพกราฟิกเพื่อนำเสนอข้อมูล อำนวยความสะดวกให้กรรมการหลักสูตรในการวางแผนการทำงาน
3. เพื่อนำข้อมูลผลการเรียนของนักศึกษาในอดีต มาพัฒนาเป็นระบบแนะนำและวางแผนการเรียนตัวของนักศึกษาได้
4. เพื่อนำข้อมูลการพยากรณ์อาชีพในอนาคตของนักศึกษาในสถาบันมาแสดงเป็นแผนภาพกราฟิกในการอำนวยความสะดวกให้หน่วยงานภายนอกได้รับทราบว่าหลักสูตรปัจจุบันของสถาบันสามารถผลิตบุคลากรที่มีความชำนาญด้านใดได้บ้าง

### 3. ทฤษฎีที่เกี่ยวข้อง (Theoretical Background)

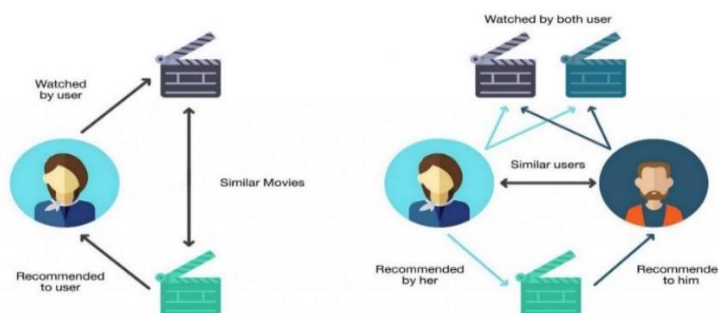
เนื่องจากโครงงาน Recommendation system for CE Curriculum administrators นี้ เป็นโครงงานที่นำความรู้ด้าน Data analytics มาประยุกต์ใช้เพื่อแก้ปัญหาที่กล่าวมาข้างต้น ทฤษฎีที่เกี่ยวข้องกับโครงงานจึงเป็นความรู้ในด้าน Data analytics ได้แก่ Recommendation system, Classification and Prediction, Visualization, และ Scikit-learn ซึ่งมีรายละเอียด ดังนี้

#### 3.1. Recommendation System

Recommendation System เป็นระบบที่จะแนะนำสิ่ง (item) ที่ “เหมาะสม” ให้กับผู้ใช้ โดย item เป็นได้ตั้งแต่ ข่าว, เนื้อหา, เพลง, course เรียน, ไปจนถึงสินค้าที่ขายในร้าน online โดยระบบสามารถให้คำแนะนำที่ต่างกันตามความชอบของแต่ละผู้ใช้ได้ [1]

ข้อควรคำนึงในการทำ recommendation system

- Relevance: นำเสนอของที่คิดว่าผู้ใช้น่าจะสนใจ
- Diversity: สิ่งที่ต้องการจะนำเสนอ นั้นจะต้องคำนึงถึงความหลากหลาย
- Novelty: นำเสนอของที่ผู้ใช้อย่างไม่เคยเห็นหรือมีความแตกต่างจากสิ่งที่ผู้ใช้ได้เห็นมาในอดีต
- Serendipity: มีทั้งความ relevant + novelty และสร้างความพึงพอใจให้ผู้ใช้



รูปที่ 3.1 ตัวอย่าง Recommendation System<sup>1</sup>

ประเภทของ Recommendation System

- 1) Content-based Filtering เป็นรูปแบบโมเดลที่แนะนำลักษณะของตัวบริการหรือสินค้าเป็นตัวตั้ง แล้วแนะนำสินค้าและบริการอื่นที่มีคุณลักษณะคล้าย ๆ กัน
- 2) Collaborative Filtering เป็นรูปแบบโมเดลที่เรียนรู้จากพฤติกรรมของผู้ใช้กับผู้ใช้คนอื่น ๆ ที่คล้ายคลึงกัน
  - Memory-based ซึ่งจะเน้นไปในทางที่ดูข้อมูลแล้วหาความสัมพันธ์ ระหว่างผู้ใช้หรือสินค้าจากข้อมูลตรง ๆ
  - Model-based ใช้เทคนิคของ machine learning เพื่อหา user embedding และ item embedding มาทำนาย rating ที่ผู้ใช้จะให้กับสินค้า หรือ relevance score

<sup>1</sup> ที่มารูปภาพ: <https://www.ldaworld.com/recommendation-system-lda/>

- Hybrid ใช้หลาย ๆ วิธีการมารวมกัน

3) Hybrid system เป็นการมัตรวมทั้งสองระหว่าง Content-based Filtering และ Collaborative Filtering เพื่อทำให้ระบบการแนะนำสมบูรณ์ขึ้น

### Memory-based

ตัวอย่างของการใช้ memory-based เช่น การใช้ nearest neighbor คือการคาดเดา rating จากการดูจากผู้ใช้ (user-based) หรือ สินค้า (item-based) ที่ใกล้เคียงกัน

โดยวิธีการของ memory-based นั้น จะใช้หลักการของ nearest neighbor คือ เก็บข้อมูลการชอบของผู้ใช้ทั้งหมดมา แล้วดูว่ามีผู้ใช้คนใด หรือ สินค้าใด ที่มีลักษณะคล้ายกันบ้างจากข้อมูลนั้น

ตัวอย่างเช่น ถ้ามีตารางที่บอกว่าผู้ใช้แต่ละคนให้คะแนน rating กับสินค้า (สมมติว่าเป็นโรงแรม) แต่ละอย่างเท่าไรบ้าง โดยจะมีตารางดังนี้ [2]

|   | 1    | 2    | 3    | 4    | 5    |
|---|------|------|------|------|------|
|   |      |      |      |      |      |
| A | 7.00 | 6.00 | 6.80 |      | 8.00 |
| B | 8.30 | 6.00 | 8.00 | 7.40 | 7.00 |
| C |      | 7.00 |      | 8.00 | 9.50 |
| D | 8.00 | 6.00 | 7.75 | ?    | 7.50 |
| E |      | 6.50 | 5.50 | 7.75 |      |

รูปที่ 3.2 Users' Ratings และ เปรียบเทียบการใช้คะแนน rating ของผู้ใช้

$$B = [8.3, 6.0, 8.0, 7.4, 4.0] \text{ และ } D = [8.0, 6.0, 7.75, ?, 7.5]^2$$

สำหรับช่องว่างนั้นคือไม่มีข้อมูลของการให้คะแนน rating ของผู้ใช้ต่อสินค้านั้น ถ้าต้องการจะเดาว่า ผู้ใช้ D จะให้คะแนนโรงแรมหมายเลข 4 เท่าไร ตามวิธีการของ user-based สามารถทำได้โดยหาผู้ใช้ที่ใกล้เคียงกับผู้ใช้ D มากที่สุดมา k คน แล้วลองเดาจาก k คนนี้ว่าผู้ใช้ D จะให้คะแนนเท่าไรอย่างสมมติว่า ถ้ากำหนดให้  $k = 1$  (คือ หาคคนใกล้เคียงที่สุดมาแค่คนเดียว) ผู้ใช้ D นั้นก็จะใกล้เคียงกับผู้ใช้ B ที่สุด เพราะต่างก็ให้คะแนนโรงแรมหมายเลข 1 และ 3 ค่อนข้างสูงด้วยกันทั้งคู่เมื่อเทียบกับโรงแรมหมายเลข 2 และ 5 ดังนั้นผู้ใช้ D น่าจะให้คะแนนกับโรงแรมหมายเลข 4 แบบประมาณกลาง ๆ เช่นเดียวกับผู้ใช้ B

จะเห็นว่าการทำ recommendation ด้วยวิธีการนี้ใช้แค่การกำหนดฟังก์ชันที่คำนวณความเหมือนของผู้ใช้หรือสินค้า (similarity function) และการคำนวณหาสิ่งที่คล้ายกันด้วยวิธีการของ nearest neighbor

<sup>2</sup> ที่มารูปภาพ: <https://medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>.

### 3.2. Classification and Prediction

เทคนิคการจำแนกประเภทข้อมูล (Data Classification) เป็นเทคนิคหนึ่งที่สำคัญของการสืบค้นความรู้บนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database: KDD) หรือ ดาต้าไมน์นิ่ง (Data Mining) เทคนิคการจำแนกประเภทข้อมูลเป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้จากกลุ่มตัวอย่างข้อมูลที่เรียกว่าข้อมูลสอนระบบ (training data) ที่แต่ละแถวของข้อมูลประกอบด้วยฟิลด์ (field) หรือแอทริบิวต์ (attribute) จำนวนมาก [3]

ประเภทของปัญหาด้าน Classification

#### 1) Binary classification (การจำแนกแบบไบนารี)

เปรียบเทียบให้เข้าใจง่ายที่สุดก็คือ ตัวแปรที่แบ่งเป็นเพียงสองหมวดหมู่ เช่น ผลลัพธ์แบบ ใช่ หรือ ไม่ใช่ ชั่ว หรือ ไม่ชั่ว หากเปรียบเทียบเป็นตัวเลขก็คือ 0 กับ 1 อัลกอริทึมที่ใช้คู่กับการจำแนกแบบไบนารีก็มี k-Nearest Neighbors Decision Trees หรือ Naive Bayes

#### 2) Multi-Class Classification (การจำแนกประเภทหลายคลาส)

ต่างกับการจำแนกแบบไบนารีตรงที่มีหมวดหมู่มากกว่าสอง ยกตัวอย่างเช่น รูปภาพที่คล้ายคลึงกับรูปภาพที่อัปโหลดเพื่อค้นหาคำศัพท์ที่คาดว่าจะพิมพ์ใน predictive keyboard คือผลลัพธ์ที่มีได้มากกว่า 2 หมวดหมู่ อัลกอริทึมที่ใช้คู่กับการจำแนกประเภทหลายคลาสก็สามารถใช้อัลกอริทึมคล้ายกับที่ใช้กับการจำแนกแบบไบนารีได้

#### 3) Multi-Label Classification (การจำแนกประเภทหลายเลเบล)

เพื่อเปรียบเทียบให้เข้าใจง่ายขึ้น ขอยกตัวอย่างเช่น รูปภาพรูปหนึ่งสามารถมีรูปดอกไม้ ทั้งฟ้า ก้อนเมฆได้ แต่รูปภาพรูปนั้นจะจัดว่าเป็นหมวดหมู่รูปวาด รูปถ่าย หรือรูปเสีย Multi-Label Classification ก็คือการเล-เบล หรือติดฉลากว่าในรูปนั้น ๆ มีดอกไม้หรือเปล่ามีก้อนเมฆหรือไม่ส่วน Multi-Class Classification จะจำแนกว่ารูปนั้นเป็นรูปวาดรูปถ่ายรูปเสีย

#### 4) Imbalanced Classification (การจำแนกแบบข้อมูลไม่เท่าเทียม)

เป็นปัญหาที่เกิดจากข้อมูลที่ไม่เท่าเทียม (Imbalanced dataset) ตัวอย่างเช่นข้อมูลการทุจริต โดยข้อมูลส่วนใหญ่ย่อมเป็นข้อมูลที่จัดว่า “ไม่ทุจริต” และจะมีเปอร์เซ็นต์น้อยที่จัดว่าเป็น “ทุจริต” เป็นต้น

การจัดเตรียมข้อมูลสำหรับการจำแนกและการทำนายข้อมูล

- 1) การทำความสะอาดข้อมูล (Data cleansing) จะเกี่ยวข้องกับการประมวลผลข้อมูลเบื้องต้นที่จะลบหรือลดข้อมูลที่มีสิ่งรบกวน (noise)
- 2) ความเกี่ยวเนื่องของข้อมูล (Relevance analysis) จะทำการตรวจสอบข้อมูลแอทริบิวต์ต่าง ๆ ว่ามีความเกี่ยวเนื่องหรือซ้ำซ้อนกันมากน้อยเพียงใด ซึ่งโดยปกติของชุดข้อมูลจะมีแอทริบิวต์ (attribute) ที่ซ้ำซ้อนกัน เพื่อที่จะหลีกเลี่ยงความซ้ำซ้อน
- 3) การเปลี่ยนแปลง/เปลี่ยนรูปข้อมูลและการลดจำนวนข้อมูล (Data transformation and reduction) ข้อมูลที่เป็นอินพุต (Input) อาจมีช่วงของข้อมูลหรือค่าของข้อมูลที่มีระยะห่างค่อนข้างมาก ดังนั้น เราอาจทำการเปลี่ยนแปลง/เปลี่ยนรูปด้วยวิธีการ normalization ที่จะทำการปรับเปลี่ยนค่าในแอทริบิวต์ (attribute) หนึ่ง ๆ ให้อยู่ในช่วงที่กำหนด

### 3.2.1. Decision Tree

เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูล

ในการเรียนรู้ของเครื่อง (machine learning) Decision Tree เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

Decision Tree ที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่า Decision Tree แบบจำแนก (classification trees) และ Decision Tree ที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่า Decision Tree แบบถดถอย (regression trees)

#### ลักษณะของ Decision Tree

Decision Tree จะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของ Decision Tree คือตัวแปร (attribute) ต่าง ๆ ของชุดข้อมูล เช่น หากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณาคือ ทักษะสภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์ Decision Tree คือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของตัวแปรค่าของตัวแปร (attribute-value pair) เช่น ทักษะสภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็น ใช่ กับ ไม่ใช่ เป็นต้น การทำนายประเภทด้วย Decision Tree จะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่า เพื่อไปยังบัพลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอบัพใบซึ่งจะแสดงผลการทำนาย

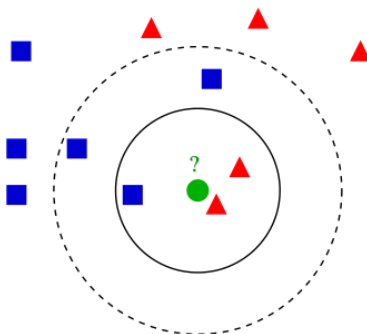
### 3.2.2. K-Nearest Neighbors (KNN)

ขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด (อังกฤษ: K-Nearest Neighbors Algorithm) เป็นวิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน (“K” ในขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่าง ๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด

กำหนดให้จุดที่พิจารณาคือ วงกลมสีเขียว ควรจัดกลุ่มให้จุดที่สนใจไปอยู่ใน คลาสแรกของสี่เหลี่ยมสีน้ำเงิน หรือ คลาสสองของสามเหลี่ยมสีแดง

ถ้า  $k=3$  แล้ว วงกลมสีเขียวจะอยู่ในคลาสสอง เพราะมี สี่เหลี่ยม 1 รูป และ สามเหลี่ยม 2 รูป อยู่ในวงกลมวงใน

ถ้า  $k=5$  แล้ว วงกลมสีเขียวจะอยู่ในคลาสแรก เพราะมี สี่เหลี่ยม 3 รูป และ สามเหลี่ยม 2 รูป อยู่ในวงกลมวงนอก



รูปที่ 3.3 ตัวอย่างการจัดกลุ่มข้อมูลของขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด<sup>3</sup>

<sup>3</sup> ที่มารูปภาพ: [https://th.wikipedia.org/wiki/ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด\\_k\\_ตัว](https://th.wikipedia.org/wiki/ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้ที่สุด_k_ตัว)

## ขั้นตอนวิธี

การนำเทคนิคของขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุดไปใช้นั้น เป็นการหาระยะห่างระหว่างแต่ละตัวแปร(Attribute) ในข้อมูล จากนั้นก็คำนวณค่าออกมา ซึ่งวิธีนี้จะเหมาะสมสำหรับข้อมูลแบบตัวเลข แต่ตัวแปรที่เป็นค่าแบบไม่ต่อเนื่องนั้นก็สามารถทำได้ เพียงแต่ต้องการการจัดการแบบพิเศษเพิ่มขึ้น อย่างเช่น ถ้าเป็นเรื่องของสี เราจะใช้อะไรวัดความแตกต่างระหว่างสีน้ำเงินกับสีเขียว ต่อจากนั้นเราต้องมีวิธีการรวมค่าระยะห่างของ Attribute ทุกค่าที่วัดมาได้ เมื่อสามารถคำนวณระยะห่างระหว่างเงื่อนไขหรือกรณีต่าง ๆ ได้ จากนั้นก็เลือกชุดของเงื่อนไขที่ใช้จัดคลาส มาเป็นฐานสำหรับการจัดคลาสในเงื่อนไขใหม่ ๆ ได้ แล้วเราจะตัดสินใจได้ว่าขอบเขตของจุดข้างเคียงที่ควรเป็นนั้น ควรมีขนาดใหญ่เท่าไร และอาจมีการตัดสินใจได้ด้วยว่าจะนับจำนวนจุดข้างเคียงตัวมันได้อย่างไร โดยขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุดมีขั้นตอนโดยสรุป ดังนี้

- 1) กำหนดขนาดของ K (ควรกำหนดให้เป็นเลขคี่)
- 2) คำนวณระยะห่าง (Distance) ของข้อมูลที่ต้องการศึกษากับกลุ่มข้อมูลตัวอย่าง
- 3) จัดเรียงลำดับของระยะห่าง และเลือกพิจารณาชุดข้อมูลที่ใกล้จุดที่ต้องการพิจารณาตามจำนวน K ที่กำหนดไว้
- 4) พิจารณาข้อมูลจำนวน k ชุด และสังเกตว่ากลุ่ม (class) ไหนที่ใกล้จุดที่พิจารณาเป็นจำนวนมากที่สุด
- 5) กำหนด class ให้กับจุดที่พิจารณา (class) ที่ใกล้จุดพิจารณามากที่สุด

## การดำเนินการหลัก

ฟังก์ชันระยะทาง(Distance Function) เป็นการคำนวณค่าระยะห่างระหว่างสองเรคคอร์ด เพื่อที่จะมาวัดความคล้ายคลึงกันของข้อมูล

ฟังก์ชันการแจกแจง (Combination Function) เป็นการรวมกันของผลลัพธ์ที่ได้จากการคำนวณค่าระยะห่าง (Distance) โดยทำการเรียงลำดับค่าระยะห่าง (Distance) จากน้อยไปมาก หลังจากนั้นดูจากค่า “K” ว่ากำหนดเป็นเท่าไร แล้วนำลำดับที่เรียงได้มาเทียบกับคลาสข้อมูลที่เรียงแล้วนำมาตอบ

## คุณสมบัติของฟังก์ชันระยะทาง (Distance Function)

- ค่าระยะทาง(ความห่าง) ที่คำนวณได้ต้องไม่ติดลบ
- ถ้าตำแหน่งเดียวกัน ฟังก์ชันต้องเป็นศูนย์(ค่าเหมือนกัน)
- การคำนวณวัดระยะทางไปกลับต้องเท่ากัน



การคำนวณค่าฟังก์ชันระยะทาง

- ใส่ค่าสัมบูรณ์ (Absolute) ให้กับค่าระยะห่าง :  $|A-B|$
- ยกกำลังสองให้กับค่าระยะห่าง :  $(A-B)^2$
- ทำการปรับให้เป็นค่ามาตรฐาน:  $|(A-\text{mean})/(SD)-(B-\text{mean})/(SD)|$

การรวมค่าระยะทาง (Distance) ในเรคคอร์ด (Record)

การวัดระยะแบบแมนฮัตตัน (Manhattan distance) เป็นการนำค่าที่คำนวณได้ในหนึ่งเรคคอร์ด (Record) มารวมกัน

ระยะทางแบบยูคลิด (Euclidean distance) เป็นการหารากที่สอง (Square Root) ในแต่ละตัวแปร(attribute) แล้วนำมาบวกกัน แล้วนำค่าที่คำนวณได้ในหนึ่งเรคคอร์ด (Record) มารวมกัน

### 3.2.3. Naive Bayesian Classifier

ตัวจำแนกแบบเบย์อย่างง่าย (Naive Bayesian Classifier) คือโมเดลการจำแนกประเภทข้อมูลที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes' Theorem และสมมติฐานที่ให้การเกิดของเหตุการณ์ต่าง ๆ เป็นอิสระต่อกัน (Independence) กำหนดให้  $P(h)$  ความน่าจะเป็นที่จะเกิดเหตุการณ์  $h$  และ  $P(h|D)$  คือความน่าจะเป็นที่จะเกิด

เหตุการณ์  $h$  เมื่อเกิดเหตุการณ์  $D$  จากตัวแปรที่กำหนดและแนวคิดของ Bayes' Theorem นั้น เราสามารถทำนายเหตุการณ์ที่พิจารณาได้จากการเกิดของเหตุการณ์ต่าง ๆ ได้ดังสมการ

$$P(h|D) = [P(D|h) * P(h)] / P(D)$$

### 3.3. Visualization

ในทศวรรษที่ผ่านมา ข้อมูลมหาศาลถูกสร้างขึ้นโดยอัตโนมัติ และเพื่อให้ได้มาซึ่ง ข้อมูลเชิงลึก (Insight) จากข้อมูลเหล่านั้น รูปแบบและวิธีการวิเคราะห์ข้อมูลต่าง ๆ ได้เกิดขึ้นจำนวนมากและหลากหลาย อย่างไรก็ตามความสามารถทางสัมผัสและสติปัญญาของมนุษย์ที่จะเข้าใจธรรมชาติของข้อมูลที่ซับซ้อนเหล่านั้นมีอยู่จำกัด จะมีหนทางหรือวิธีการใดที่ทำให้มนุษย์สามารถเข้าใจข้อมูลได้อย่างง่ายและรวดเร็วได้บ้าง

การวิเคราะห์ด้วยภาพ คือการใช้เหตุผลเชิงวิเคราะห์ (Analytical Reasoning) ซึ่งว่าด้วยการมองดูข้อมูลในเชิงปริมาณหรือปริมาตรแล้วสามารถสกัดเอาความรู้หรือข้อสรุปออกมาได้ ประกอบด้วยหลายศาสตร์มากมาย เช่น Cognitive Science, Information Systems, Graphic & Interactive Design, Human-Computer Interaction, Mathematics & Statistical Methods, Data Analysis ซึ่งมี Interactive Visual Interfaces เป็นเครื่องมือและสื่อกลาง [4]



รูปที่ 3.4 ตัวอย่าง Visual Analytics สรุปการขาย<sup>4</sup>

การวิเคราะห์ด้วยภาพเป็นการวิเคราะห์ข้อมูลโดยใช้รูปภาพเพื่อสื่อความ ไม่มีส่วนเกี่ยวข้องและเป็นคนละเรื่องกับศาสตร์ของการประมวลผลภาพดิจิทัล หรือ Digital Image Processing

Visual Analytics จะช่วยเพิ่มกำลังความสามารถในการรับรู้และเข้าใจของมนุษย์ได้ต่อไปนี้

- ลดการทำงานในส่วนความจำ โดยใช้ภาพที่มองเห็นด้วยตาเปล่าเป็นสื่อรับรู้
- ลดการค้นหาที่หนักและซับซ้อน โดยการนำเสนอข้อมูลที่มีขนาดใหญ่ในพื้นที่แสดงขนาดเล็ก
- ใช้รูปแบบเฉพาะเพื่อให้เราสามารถเข้าใจความหมายได้อย่างรวดเร็ว เช่น แผนภูมิวงกลม (Pie chart) หมายถึงสัดส่วนของข้อมูล
- ช่วยในการอนุมานความสัมพันธ์ให้ง่ายขึ้น
- ติดตามเหตุการณ์หรือปรากฏการณ์ในข้อมูลหลายเรื่อง ๆ ให้สามารถมองเห็นได้ด้วยตาเปล่า
- มีเครื่องมือเพื่อให้จัดการเปลี่ยนแปลงค่าพารามิเตอร์ต่าง ๆ ทำให้ทำการสำรวจได้มากยิ่งขึ้น ต่างจากรูปภาพนิ่งที่ไม่สามารถเปลี่ยนแปลงค่าใด ๆ ได้

<sup>4</sup> ที่มารูปภาพ: <https://bigdata.go.th/big-data-101/visual-analytics/>

### 3.4. Scikit-learn

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

นอกจากนี้ Scikit-learn ยังเป็น Open Source ที่เปิดให้ผู้ที่สนใจสามารถเข้าไปพัฒนาต่อยอดได้ และที่ทำให้ทุกคนต่างยอมรับคือเป็นแหล่งรวม Library และอัลกอริทึมที่เน้นไปในด้านของ การเรียนรู้ของเครื่อง (Machine Learning) ซึ่งมีส่วนในการทำ แบบจำลองข้อมูล (Data Modeling) ทั้งนี้ อีกหนึ่งปัจจัยที่ทำให้มีผู้ใช้เยอะ และเป็นเครื่องมือที่แนะนำสำหรับมือใหม่ เพราะเป็น Interface ระดับสูง ให้มือใหม่สามารถเข้าใจภาพรวมและ ขั้นตอนการทำงาน ของการเรียนรู้ของเครื่อง (Machine Learning) ได้เครื่องมือที่ผู้ใช้งานสามารถนำไปใช้ในโปรเจกต์ได้ [5]

#### 1) Classification

หนึ่งในตัวอย่างของ Classification คือการแยกอีเมลว่าเป็นสแปมหรือไม่ Classification ถือว่าเป็นหนึ่งในแขนงของ Supervised Learning (การเรียนรู้แบบมีผู้สอนนั้น คือการทำให้คอมพิวเตอร์สามารถหาคำตอบของปัญหาได้ด้วยตัวเอง หลังจากเรียนรู้จากชุดข้อมูลตัวอย่างไปแล้วระยะหนึ่ง) การเรียนรู้ของอัลกอริทึมจากชุดข้อมูล (Datasets) ที่มีคำตอบที่ถูกต้อง

#### 2) Regression

เป็นเครื่องมือเพื่อเข้าใจความสัมพันธ์ระหว่างข้อมูล Input กับ Output ซึ่งก็ถือว่าเป็นอีกหนึ่งแขนงของ Supervised Learning เช่นกัน ตัวอย่างของอัลกอริทึมนี้คือการเข้าใจพฤติกรรมของราคาหุ้น

#### 3) Clustering

คือการจำแนกกลุ่มข้อมูลที่มีคุณสมบัติคล้ายกัน เครื่องมือนี้ถือว่าเป็นแขนงของ Unsupervised Learning (เป็นการเรียนรู้ที่ไม่มีผู้สอน โดยที่เครื่องจักรนั้น ๆ จะทำการเรียนรู้สิ่งต่าง ๆ ด้วยตัวเอง โดยไม่ต้องตั้งค่าเป้าหมายของแต่ละข้อมูล) และตัวอย่างของการใช้อัลกอริทึมนี้คือการจำแนกกลุ่มลูกค้าตามพื้นที่

#### 4) Model selection

อัลกอริทึมเพื่อใช้เปรียบเทียบ ตรวจสอบ และเลือกโมเดล พารามิเตอร์ ที่เหมาะสมกับชุดข้อมูลที่สุดในโปรเจกต์ จะช่วยเพิ่มความแม่นยำของอัลกอริทึม การเรียนรู้ของเครื่อง (Machine Learning) ได้

#### 5) Pre-processing

ในขั้นตอนของ Data Analysis หรือการเข้าใจและวิเคราะห์ข้อมูล อาจต้องมีการแก้ไขให้ข้อมูลอยู่ในรูปแบบที่เรานำไปใช้งานได้ เครื่องมือตัวนี้ของ Scikit-learn จะสามารถช่วยจัดการกับข้อมูลได้

#### 4. งานวิจัยที่เกี่ยวข้อง (Related Works)

##### 4.1. วิเคราะห์และพัฒนาระบบแนะนำหนังสือคอมพิวเตอร์ แบบออนไลน์โดยใช้เทคนิคการกรองแบบอิงเนื้อหา Development and Analysis of an Online Computer Books Recommendation System Using Content-Based Filtering [6]

ทำการวิจัยเพื่อ 1) พัฒนาระบบแนะนำหนังสือคอมพิวเตอร์แบบออนไลน์โดยใช้เทคนิคการกรองแบบอิงเนื้อหา 2) เพื่อประเมินความพึงพอใจของการพัฒนาระบบแนะนำหนังสือคอมพิวเตอร์แบบออนไลน์ที่พัฒนาขึ้น โดยนำชุดข้อมูลที่มีมาผ่านขั้นตอน การกรองแบบอิงเนื้อหา 1) การสร้างเมตริกซ์ของผู้ใช้ชั้นข้อมูลจริง 2) ทำนายเรตติ้งด้วยเทคนิคการกรองแบบอิงเนื้อหา 3) สร้างเมตริกซ์ผู้ใช้ชั้นข้อมูลเทียมผ่าน อัลกอริทึม Naïve Bayes พบว่า 1) ความสามารถของระบบจะขึ้นอยู่กับจำนวนของเรตติ้งที่ให้ไว้ โดยถ้าต่ำกว่า 5 เล่ม ผลการแนะนำจะไม่ตรงกับความต้องการ 2) เทคนิคการกรองแบบอิงเนื้อหาจะเหมาะสมกับข้อมูลหนังสือประมาณ 600 เรื่อง แต่ถ้ามีข้อมูลหนังสือมากกว่า 1000 เรื่องขึ้นไปจะทำให้ประสิทธิภาพการทำงานของระบบต่ำลง

##### 4.2. ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล A Prediction System for Undergraduate Student Dropout at Faculty of Science, Buriram Rajabhat University using Data Mining Techniques [7]

ทำการวิจัยเพื่อ 1) สร้างแบบจำลองการทำนายการพ้นสภาพของนักศึกษา 2) พัฒนาระบบทำนายการพ้นสภาพด้วยเทคนิคเหมืองข้อมูล โดยใช้คลังข้อมูล จากฐานข้อมูลนักศึกษาคณะที่มีอัตราการพ้นสภาพมากที่สุด นำมาเข้ากระบวนการเตรียมข้อมูล และจำแนกประเภทโดย อัลกอริทึม J48 และวัดประสิทธิภาพโดยวิธีการตรวจสอบแบบไขว้ พบว่า 1) Attribute GPA เนื่องด้วยตัว GPA มีข้อมูลที่หลากหลายมากซึ่งไม่เหมาะกับอัลกอริทึมการจำแนกอย่างมาก โดยแก้ไขปัญด้วยการ ปรับให้อยู่ในรูปแบบที่จำแนกได้ง่ายโดยแบ่งข้อมูลออกเป็นช่วง ๆ โดย Week =  $GPA < 1.6$  Medium =  $GPA 1.6-1.99$  Good =  $GPA 2.0-2.5$  Best =  $GPA > 2.5$

##### 4.3. การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ The Comparison of Performance Models for Predicting Students Success in Vocation Education [8]

ทำการวิจัยเพื่อ 1) ศึกษาปัจจัยที่มีผลต่อการสำเร็จการศึกษาในสถานศึกษา 2) พัฒนาแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ 3) อำนวยความสะดวกให้ครูแนะแนวและผู้บริหารในการ

วางแผนการแนะแนว โดยใช้ข้อมูลจาก สถานศึกษาอาชีวศึกษาเอกชนในจังหวัดกาฬสินธุ์ จำนวน 12 แห่ง และนำข้อมูลที่ได้มาผ่านกระบวนการแปลงข้อมูล แล้ววิเคราะห์ปัจจัยเพื่อเลือกกลุ่มของแอดทริบิวต์ โดยใช้เทคนิค Gain Ratio Attribute Evaluation แล้วนำมาสร้างแบบจำลองพยากรณ์การสำเร็จการศึกษา ผ่าน 6 เทคนิค 1) C4.5 2) Random Forest 3) Random Tree 4) REP Tree 5) K-NN 6) SVM แล้วนำมาผ่านกระบวนการวัดประสิทธิภาพ โดยใช้เทคนิคแบบ 10-fold cross validation พบว่า 1) แบบจำลองพยากรณ์ที่มีประสิทธิภาพสูงสุดคือ C4.5

## 5. ขอบเขตของโครงการ (Scope)

- ข้อมูลผลการเรียนในอดีตจะได้จากสำนักทะเบียนและประมวลผล โดยกรรมการหลักสูตรจะเป็นผู้ร้องขอข้อมูลดังกล่าวและนำเข้าระบบ
- การทำนายต่างๆ จะใช้ข้อมูลเพียง 2 แหล่งคือข้อมูลผลการเรียนของนักศึกษาจากสำนักทะเบียนและประมวลผล และแบบสอบถามการปฏิบัติงานของบัณฑิตเท่านั้น
- Algorithm ในระบบทั้งหมดจะใช้ Algorithm พื้นฐานตาม library ของ scikit-learn เท่านั้น ไม่มีการประยุกต์ใช้จาก research paper อื่นๆ

## 6. การพัฒนาโครงการ (Project Development)

### 6.1. System Diagram

ระบบรองรับการใช้งานโดยแบ่งตามประเภทผู้ใช้งานได้ 4 ประเภท ได้แก่ 1 Student (นักศึกษา) 2 Teacher (อาจารย์) 3 Committee (กรรมการหลักสูตร) 4 Guest (บุคคลภายนอก) ซึ่งสามารถแบ่งการทำงานได้ดังนี้

#### ส่วนหน้าเว็บแอปพลิเคชัน

1. Download CSV Template คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการดาวน์โหลดตัวอย่าง file CSV เพื่อให้ผู้ใช้งานปรับใช้ file ของตน หรือ แก้ไข file ที่ได้มาตามรูปแบบที่ถูกต้อง
2. Analytic Student คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการนำข้อมูลของผู้ใช้งานไปประมวลผล
3. Prediction Grade Interface คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อแสดงผลการประมวลผลของผู้ใช้งานในส่วน Grade(เกรด)
4. Prediction career Interface คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อแสดงผลการประมวลผลของผู้ใช้งานในส่วน Career(อาชีพ)

5. Profession GUI คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อแสดงผลการประมวลผลของผู้ใช้งานในส่วนคะแนนความถนัดในรูปแบบของกราฟ
6. Student In Class คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการแสดงผลข้อมูลนักศึกษาในวิชานั้น
7. Prediction Student In Class คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการแสดงผลผลการพยากรณ์ของนักศึกษาในวิชานั้น ๆ
8. Is Retire / Is Quit Out คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการแสดงผลผลการพยากรณ์การลาออกและถูกให้ออกของนักศึกษาทั้งหมด
9. Input Data / Update Data คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการเพิ่มข้อมูลหรือแก้ไขข้อมูลนักศึกษา
10. Overall Predict Student Career Outcome คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อแสดงผลผลการพยากรณ์ของนักศึกษาทั้งหมด
11. Overall Student Career Outcome คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อแสดงผลการพยากรณ์ของนักศึกษาปีการศึกษาสุดท้ายและบัณฑิตทั้งหมด
12. Upload Subject Data คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการเพิ่มข้อมูลหลักสูตร
13. Login UI คือส่วนของหน้าเว็บแอปพลิเคชันเพื่อรองรับการเข้าสู่ระบบของผู้ใช้งาน

#### ส่วน API

1. Simple CSV Generator คือส่วน API ที่รองรับการร้องขอตัวอย่าง file CSV โดยจะตอบกลับการร้องขอด้วยตัวอย่าง file CSV ที่ระบบจัดเตรียมไว้
2. Request All Prediction Handler คือส่วน API ที่รองรับการร้องขอการประมวลผลจากแบบจำลองพยากรณ์ ซึ่งจะมี parameter ได้แก่ ข้อมูลที่ผู้ใช้งานต้องการผลลัพธ์ ซึ่ง API จะตอบกลับด้วยผลลัพธ์การพยากรณ์
3. Req Student In Class คือส่วน API ที่รองรับการร้องขอ ข้อมูลของนักศึกษาในวิชานั้น ๆ ซึ่งจะมี parameter ได้แก่ รหัสวิชา และ ปีการศึกษา
4. Req Is Retire Is Quit Handler คือส่วน API ที่รองรับการร้องขอผลลัพธ์จากแบบจำลองการแนะนำ KNN และ แบบจำลองพยากรณ์ การลาออกและถูกให้ออกของนักศึกษาทั้งหมด ซึ่งมี parameter ได้แก่ ข้อมูลนักศึกษา
5. Req Predict Student In Class คือส่วน API ที่รองรับการร้องขอผลลัพธ์จากแบบจำลองพยากรณ์ของนักศึกษาในวิชานั้น ๆ ซึ่งมี parameter ได้แก่ รหัสวิชา และ ปีการศึกษา

6. Request All Outcome Predict of All Student คือส่วน API ที่รองรับการร้องขอผลลัพธ์แบบจำลองการแนะนำ KNN ของนักศึกษาทั้งหมดของวิชานั้น ๆ ซึ่งมี parameter ได้แก่ รหัสวิชา และ ปีการศึกษา
7. Recommend Handler คือส่วน API ที่รองรับการร้องขอการจัดการกับ column ของ Grade ที่ไม่มีข้อมูลผ่านการใช้แบบจำลองการแนะนำ KNN ซึ่งมี parameter ได้แก่ ข้อมูลที่ต้องการการจัดการกับ column Grade ที่ไม่สมบูรณ์
8. Subject Group คือส่วน API ที่รองรับการจัดหมวดหมู่ของวิชาทั้งหมด ซึ่งมี parameter ได้แก่ ข้อมูลที่ต้องการจัดหมวดหมู่ เช่น Software, Hardware, Network
9. Calculate overall (old) Outcome คือส่วน API ที่รองรับการร้องขอผลลัพธ์การประมวล ผลลัพธ์ของนักศึกษาปีการศึกษาสุดท้ายและบัณฑิตทั้งหมด เช่น Software Engineer 10 คน, System Administrator 10 คน
10. Occupation Group คือส่วน API ที่รองรับการจัดหมวดหมู่ของสายอาชีพจากข้อมูลที่ได้รับทั้งหมด ซึ่งมี parameter ได้แก่ ข้อมูลที่ต้องการจัดหมวดหมู่
11. Add Update Edit to Database คือส่วน API ที่รองรับ การจัดการ การเพิ่มข้อมูลของนักศึกษาลงในฐานข้อมูล ซึ่งมี parameter ได้แก่ ข้อที่ต้องการเพิ่มในฐานข้อมูล
12. Subject Add คือส่วน API ที่รองรับ การจัดการ การเพิ่มข้อมูลหลักสูตรลงในฐานข้อมูล ซึ่งมี parameter ได้แก่ ข้อที่ต้องการเพิ่มในฐานข้อมูล
13. Login Handler คือส่วน API ที่รองรับ การจัดการ การเข้าสู่ระบบของผู้ใช้งาน ซึ่งมี parameter ได้แก่ ชื่อผู้ใช้งาน และ รหัสผ่าน

#### ส่วนของฐานข้อมูล

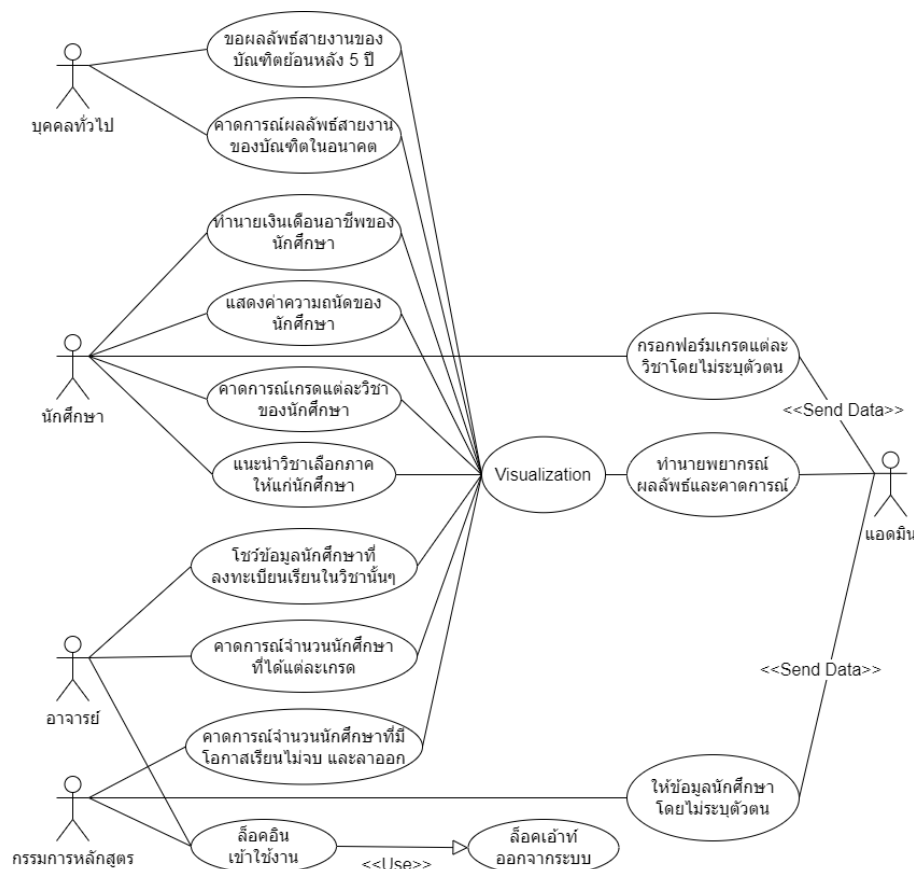
1. ข้อมูล Model (ต่อเนื่อง/นานาชาติ/ปกติ) คือข้อมูลของแบบจำลองพยากรณ์ ของหลักสูตร (ต่อเนื่อง/นานาชาติ/ปกติ) ที่ผ่านการเรียนรู้มาแล้ว
2. ข้อมูลนักศึกษา คือตารางเก็บข้อมูลของนักศึกษาของปีการศึกษาสุดท้ายและบัณฑิต**ทั้งหมด**
3. ข้อมูลหลักสูตร คือตารางเก็บข้อมูลของหลักสูตร
4. User คือตารางเก็บข้อมูลของผู้ใช้งานที่เป็นกรรมการหลักสูตรและอาจารย์

## ส่วนของ Analytics Model

1. Prediction Model ตำแหน่ง (ต่อเนื่อง/นานาชาติ/ปกติ) คือแบบจำลองพยากรณ์สำหรับการพยากรณ์ตำแหน่งอาชีพของนักศึกษา หลักสูตร (ต่อเนื่อง/นานาชาติ/ปกติ)
2. Prediction Model เงินเดือน (ต่อเนื่อง/นานาชาติ/ปกติ) คือแบบจำลองพยากรณ์สำหรับการพยากรณ์เงินเดือนของนักศึกษา หลักสูตร (ต่อเนื่อง/นานาชาติ/ปกติ)
3. Prediction Model ระยะเวลาในการหางาน (ต่อเนื่อง/นานาชาติ/ปกติ) คือแบบจำลองพยากรณ์สำหรับการพยากรณ์ระยะเวลาในการหางานของนักศึกษา หลักสูตร (ต่อเนื่อง/นานาชาติ/ปกติ)
4. Prediction Model Is Quit Retire (ต่อเนื่อง/นานาชาติ/ปกติ) คือแบบจำลองพยากรณ์สำหรับการพยากรณ์การลาออกและการถูกให้ออกของนักศึกษา หลักสูตร (ต่อเนื่อง/นานาชาติ/ปกติ)
5. KNN Model เกรด (ต่อเนื่อง/นานาชาติ/ปกติ) คือแบบจำลองการแนะนำ KNN สำหรับการพยากรณ์เกรดรายวิชาของนักศึกษา หลักสูตร (ต่อเนื่อง/นานาชาติ/ปกติ)

## 6.2. ส่วนการออกแบบ (Design)

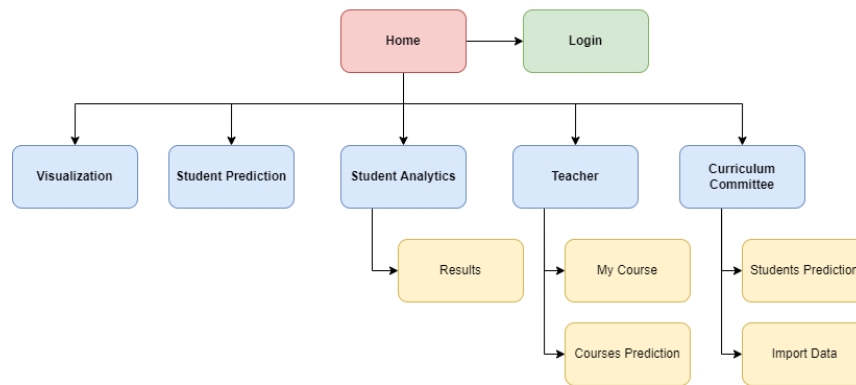
### 6.2.1. Use Case Diagram



รูปที่ 6.1 Use Case diagram



### 6.2.2. Sitemap



รูปที่ 6.2 Sitemap

หน้า Visualization แสดงหน้าหน้าสถิติบัญชีย้อนหลัง ตามภาคผนวก ภาพที่ 3

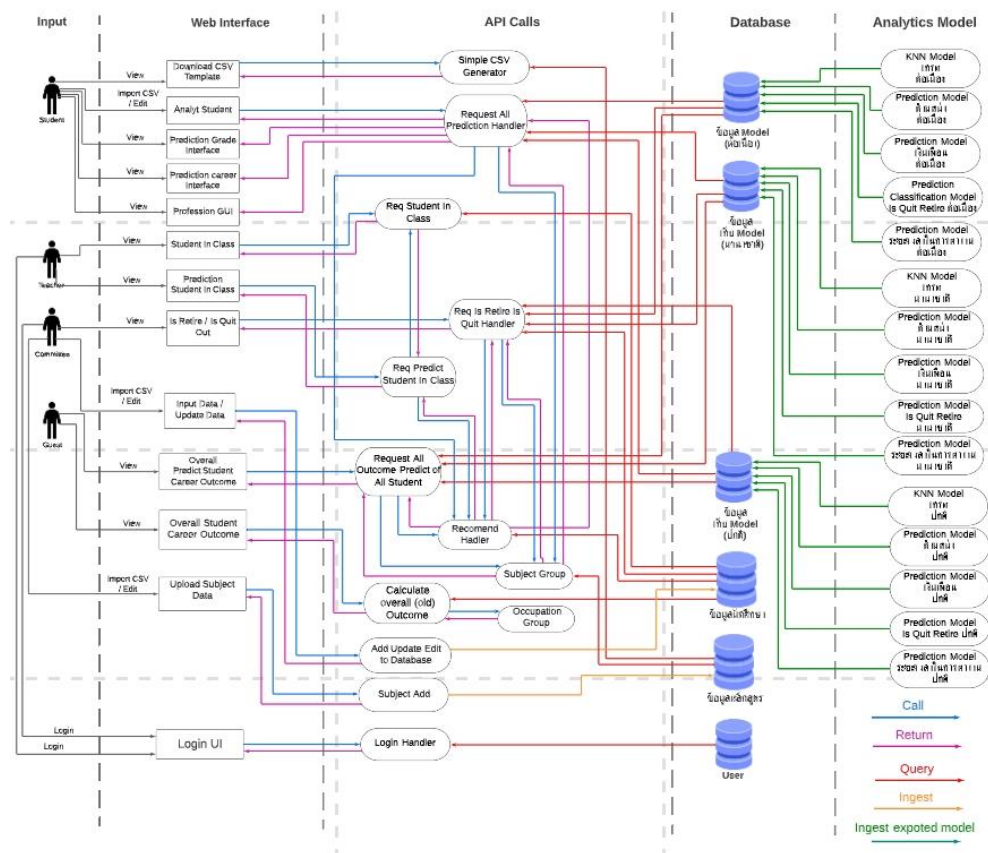
หน้า Student Prediction แสดงหน้าข้อมูลพยากรณ์นักศึกษา ตามภาคผนวก ภาพที่ 4

หน้า Student Analytics แสดง หน้าวิเคราะห์ข้อมูลนักศึกษา ตามภาคผนวก ภาพที่ 5

หน้า Curriculum Committee แสดงหน้าสำหรับ Curriculum Committee ตามภาคผนวก ภาพที่ 6

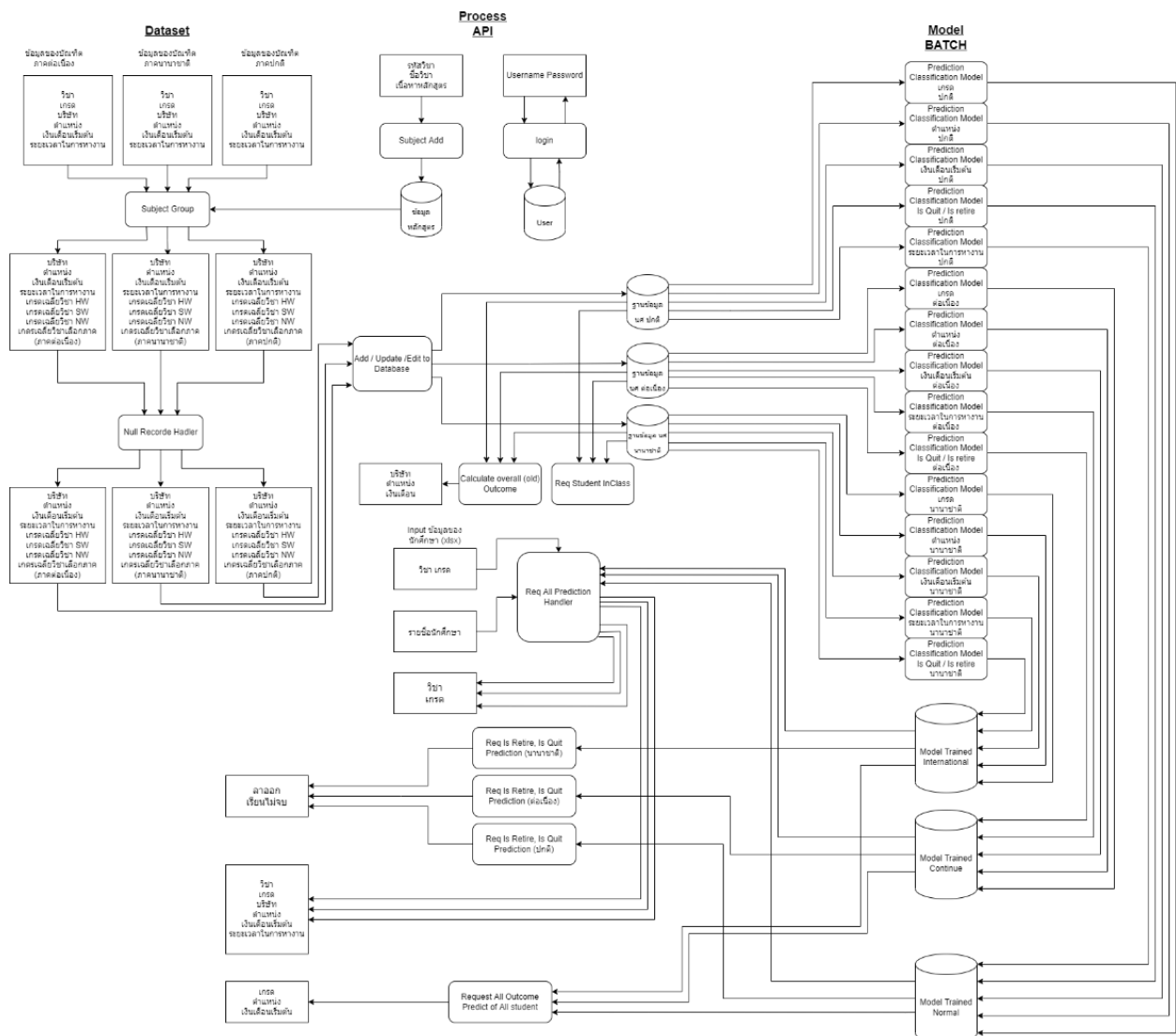
หน้า Teacher แสดงหน้าสำหรับ Teacher ตามภาคผนวก ภาพที่ 7

### 6.2.3. แผนผังรวมการทำงานของระบบ



รูปที่ 6.3 แผนผังการรวมการทำงานของโปรเจค

### 6.2.4. Data Flow Diagram

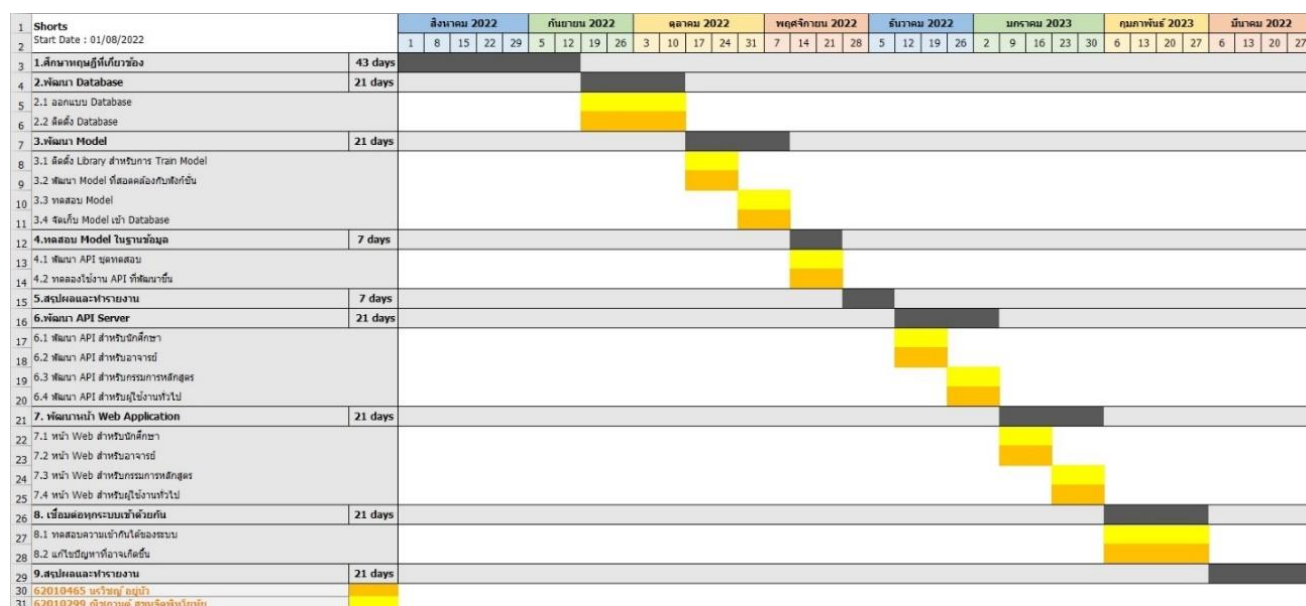


รูปที่ 6.4 Data flow and Analytic Modules

### 6.3. แนวทางการทดสอบและการวัดประสิทธิภาพ (Test and Performance Evaluation Approaches )

- 1) การทดสอบด้วย Confusion Matrix วัดประสิทธิภาพความถูกต้องของ Model โดยใช้สูตร Accuracy ( สูตร  $(TP+TN) / (TP+TN+FP+FN)$  )
- 2) การทดสอบการวัดด้วย Classification accuracy หาสัดส่วนของจำนวนความถูกต้องในการ Predict ต่อจำนวนที่นำมา Predict ทั้งหมด
- 3) การทดสอบการวัดด้วย Precision หาค่าการทำนายผิด positives (ทำนายผิดแต่ไม่เป็นผลเสีย)
- 4) การทดสอบการวัดด้วย Recall หาค่าการทำนายผิด negative แล้วเสียหาย (ทำนายผิดและมีผลเสีย)
- 5) การทดสอบการวัดด้วย F1-score โดยหาค่าเฉลี่ยระหว่าง Precision กับ Recall

## 7. แผนการดำเนินงานโครงการ (Gantt Chart)



รูปที่ 7.1 Gantt Chart

## 8. ประโยชน์ที่คาดว่าจะได้รับ (Expected Benefits)

- ได้ระบบรวบรวมข้อมูลผลการเรียนของนักศึกษาและข้อมูลแบบสำรวจการทำงานของบัณฑิต แล้วนำมาวิเคราะห์และนำเสนอข้อมูลที่เป็นประโยชน์ในการบริหารหลักสูตรของกรรมการหลักสูตร
- มีระบบที่สามารถแนะนำ วางแผน และพยากรณ์อาชีพในอนาคตจากผลการเรียนของนักศึกษา

## 9. ผลการศึกษาเทคโนโลยีที่ใช้พัฒนา

ทดลองใช้งาน Pandas : การทดลองใช้ pandas ทดสอบการทำ กระบวนการสกัด กระบวนการแปลง และ กระบวนการโหลด ข้อมูล จากผลการทดลองพบว่า เป็น library ที่ใช้งานง่าย แต่จำเป็นต้องมีความรู้และความเข้าใจเรื่องโครงสร้างของข้อมูลที่อยู่ในรูปแบบตารางเป็นอย่างดี

ทดลองใช้งาน Scikit learn : ตัวอย่างทดลองใช้งาน scikit-learn สามารถใช้ในการจดจำภาพของตัวเลขที่เขียนด้วยลายมือได้ อย่างไร ตั้งแต่ 0-9 ชุดข้อมูลตัวเลขประกอบด้วยภาพตัวเลขขนาด 8x8 พิกเซล จากนั้น จากนั้นแบ่งข้อมูลออกเป็นชุดฝึกและทดสอบชุดย่อย และใส่ตัวแยกประเภทเวกเตอร์ที่รองรับบน function train ของ scikit learn จากผลการทดลองพบว่า เป็น library ที่ใช้งานยาก เนื่องด้วยผู้ใช้งานจำเป็นต้องมีความรู้พื้นฐานด้าน การเรียนรู้ของเครื่อง (Machine Learning) เป็นอย่างดี แต่การใช้งานต่าง ๆ เช่น การเรียกการทำงานที่ตัว library มีมาให้ การจัดการกับข้อมูลที่อยู่ในรูปแบบตาราง นั้นใช้งานง่าย

ทดลองใช้งาน Django Framework : การติดตั้ง pip install Django โดยจะติดตั้งลงใน virtual environment โดยใช้งานผ่าน virtual environment wrapper เพื่อสะดวกในการจัดเก็บ file และใช้งาน จากผลการทดลองพบว่า การใช้งาน : ใช้งานง่ายเนื่องจากเป็น Architectural Style แบบ MVT ทำให้ constructure ของ project เข้าใจง่าย การ config : ค่อนข้างที่จะยุ่งยากและไม่ยืดหยุ่นเนื่องด้วยตัว framework รองรับแค่ Database แบบ SQL ถ้าอยากให้รองรับแบบ No-SQL ต้องลง library เพิ่มเติม

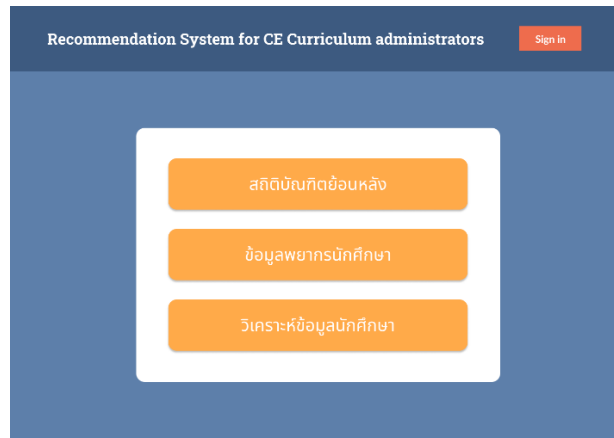
ทดลองใช้งาน MongoDB Atlas : การติดตั้งจะติดตั้ง ผ่าน Website โดยจะ monitoring Database ผ่าน Application MongoDB Compass และเชื่อมต่อกับตัว project ที่พัฒนาผ่าน connection string ที่ MongoDB เตรียมไว้ให้เลย ผลการทดลองพบว่า การใช้งาน : ใช้งานได้ง่ายผ่าน UI ของ MongoDB Compass การ Config : เข้าใจง่ายโดยปรับแต่งได้ผ่าน MongoDB Compass

## 10. เอกสารอ้างอิง (Reference)

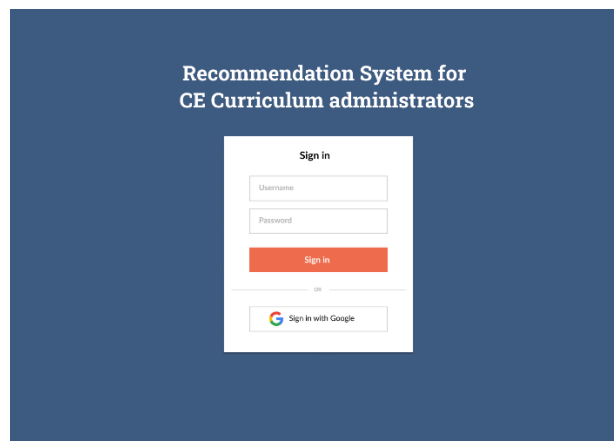
- [1] Sattawat Boonchoo. (2562). “Recommendation System ระบบผู้ช่วยแนะนำที่รู้ใจเรามากกว่าตัวเราเอง”. [ออนไลน์]. ข้อมูลจาก: <https://www.ldaworld.com/recommendation-system-lda/>
- [2] Sirinart Tangruamsub. (2562). “Recommendation System (แบบสรุปไม่ค่อยสั้น)”. [ออนไลน์]. ข้อมูลจาก: <https://medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>.
- [3] Achieve.Plus. (2563). “4 Classification ที่สำคัญใน Supervised Learning”. [ออนไลน์]. ข้อมูลจาก: <https://medium.com/achieve-space/4-classification-ที่สำคัญใน-supervised-learning-a64e75250141>.
- [4] Saksit Srimarong. (2564). “Visual Analytics การวิเคราะห์ด้วยภาพ”. [ออนไลน์]. ข้อมูลจาก: <https://bigdata.go.th/big-data-101/visual-analytics/>
- [5] Achieve.Plus. (2563). “รู้จักใช้ Scikit-Learn เหมือนมีโปรในเกม”. [ออนไลน์]. ข้อมูลจาก: <https://medium.com/achieve-space/รู้จักใช้-scikit-learn-เหมือนมีโปรในเกม-4cebd3195518>.
- [6] วลัยนุช สกุลนุ้ย. (2554). วิเคราะห์และพัฒนาระบบแนะนำหนังสือคอมพิวเตอร์แบบออนไลน์ โดยใช้เทคนิคการกรองแบบอิงเนื้อหา. ข้อมูลจาก: [http://www.rpu.ac.th/Library\\_web/doc/RC\\_RR/2554\\_ComBus\\_Walainush.pdf](http://www.rpu.ac.th/Library_web/doc/RC_RR/2554_ComBus_Walainush.pdf)
- [7] นนทวัฒน์ ทวีชาติ, อรยา เฟื่องประจัญ, วิไลรัตน์ ยาทองไชย และชูศักดิ์ ยาทองไชย. (2563). ระบบทำนายการฟื้นฟูสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล. ข้อมูลจาก: <https://ph02.tci-thaijo.org/index.php/scibru/article/view/242082>
- [8] พัฒนพงษ์ ดลรัตน์, จาริ ทองคำ. (2560). การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ. ข้อมูลจาก: [http://www.journal.msu.ac.th/upload/articles/article2154\\_27225.pdf](http://www.journal.msu.ac.th/upload/articles/article2154_27225.pdf)

## ภาคผนวก

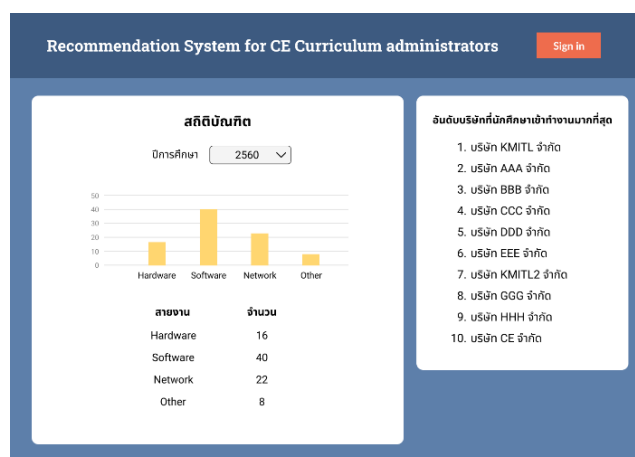
### Web Application UI



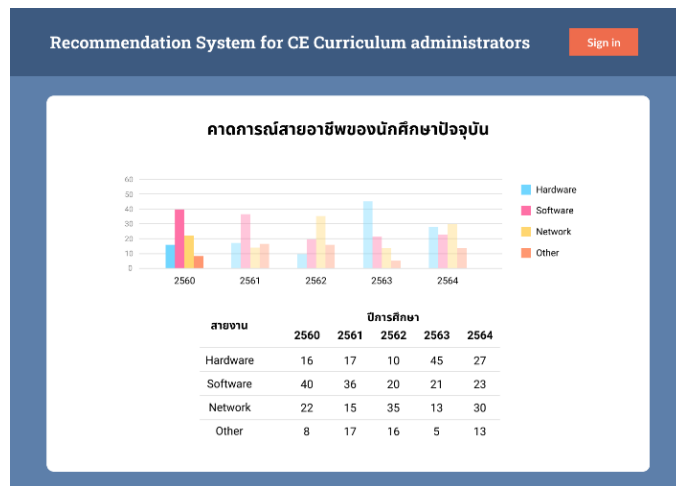
ภาพที่ 1 หน้าหลัก



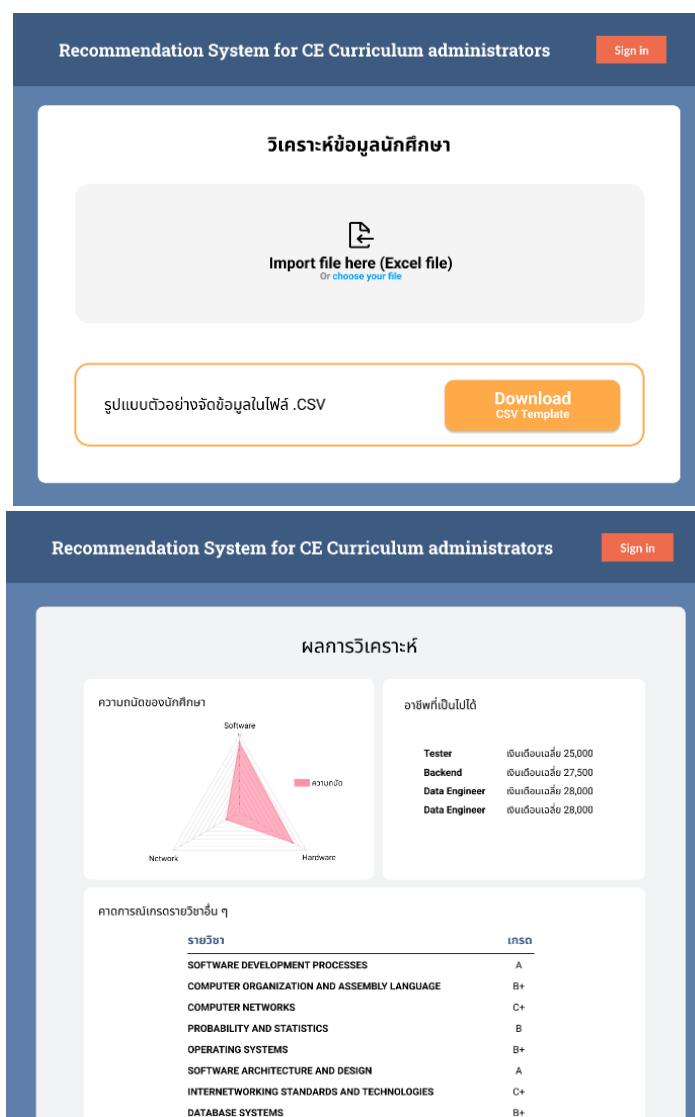
ภาพที่ 2 หน้า Login



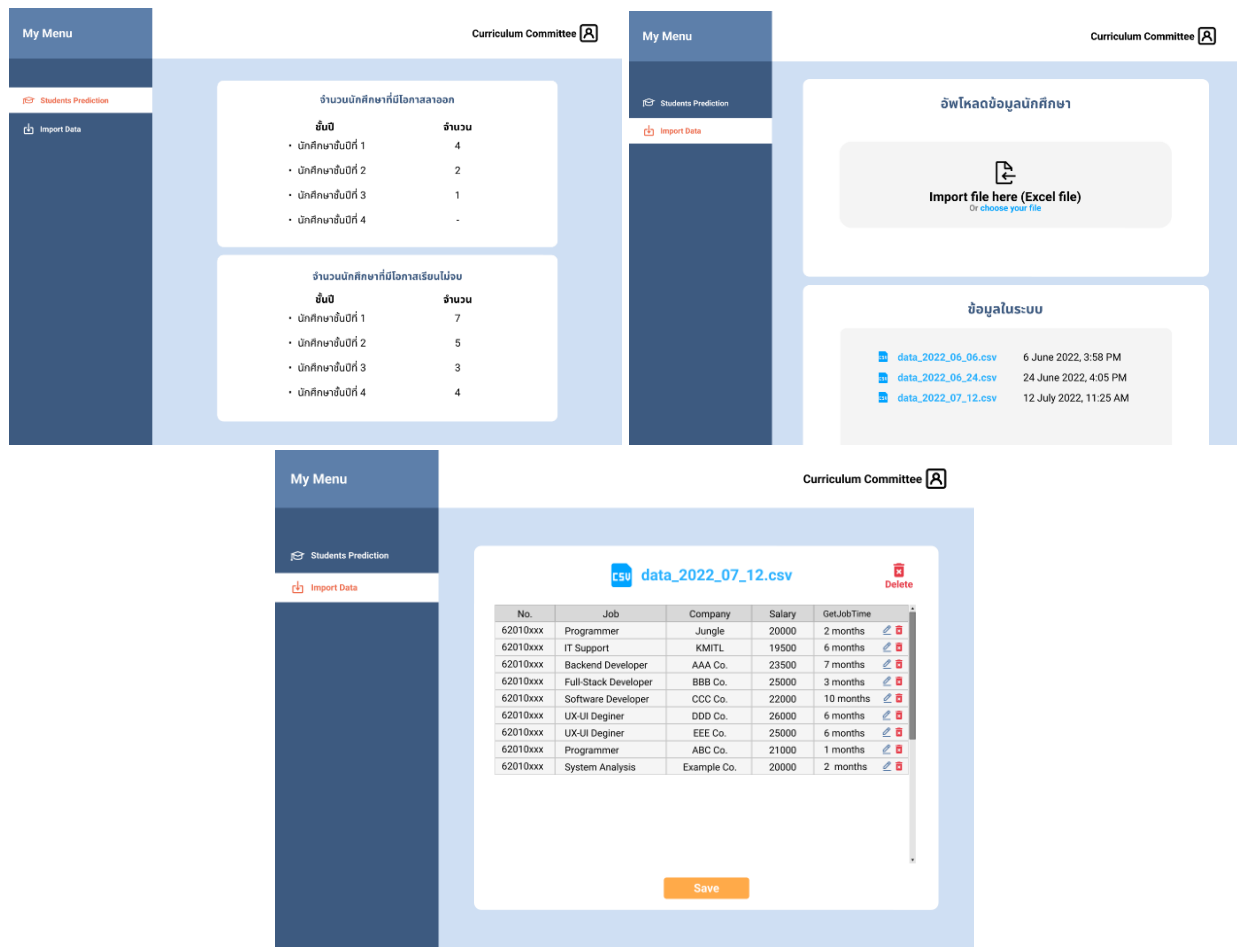
ภาพที่ 3 หน้าสถิติบัณฑิตย้อนหลัง



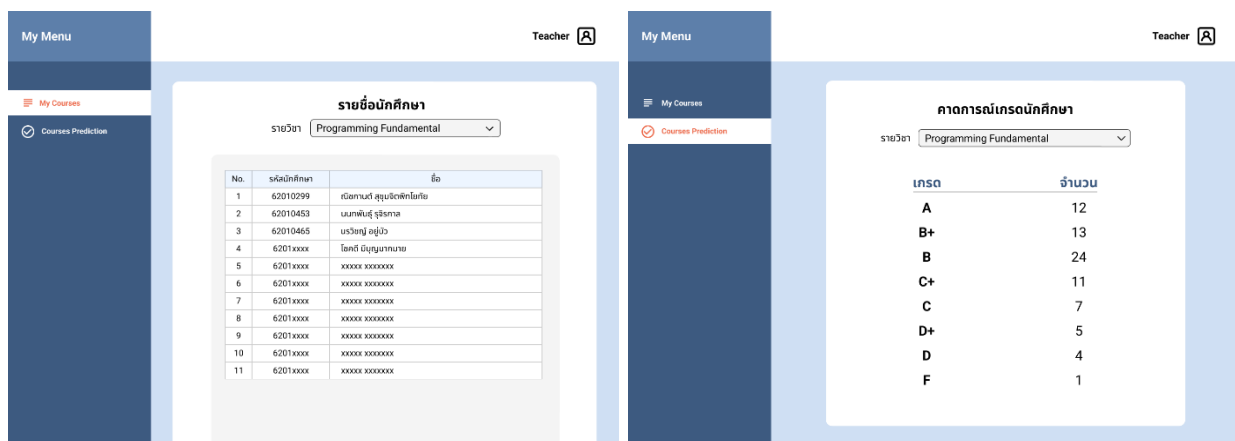
ภาพที่ 4 หน้าข้อมูลพยากรณ์นักศึกษา



ภาพที่ 5 หน้าวิเคราะห์ข้อมูลนักศึกษา



ภาพที่ 6 หน้าสำหรับ Curriculum Committee



ภาพที่ 7 หน้าสำหรับ Teacher