# รายงานความก้าวหน้าวิชา CE Project

**ครั้งที่ 4**

**ระหว่างวันที่ 24 ก.ย. 65 ถึงวันที่ 21 ต.ค. 65**

1. ชื่อโครงงาน (อังกฤษ) Analytics and Prediction System for CE Curriculum administrators

## การดำเนินงานมีความก้าวหน้า 35 % (ใช้ค่า **% Complete** จาก MS Project)

มีความก้าวหน้าเพิ่มขึ้นจากรายงานความก้าวหน้า ครั้งก่อน 7 %

□ เร็วกว่าแผน วัน □ ช้ากว่าแผน 34 วัน

## **รายละเอียดความก้าวหน้า**

หัวข้อการพัฒนาโครงงานตาม Gantt Chart

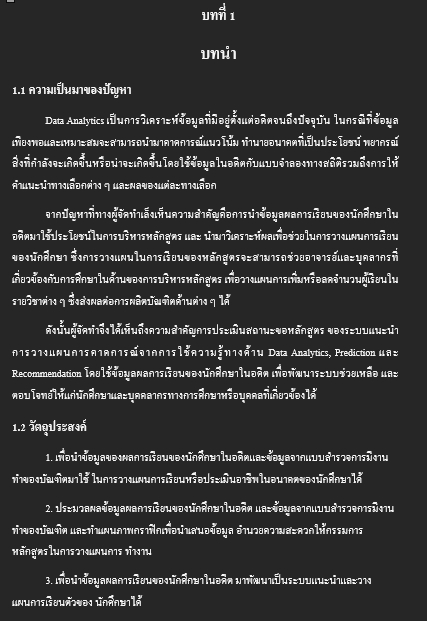
**ศึกษาทฤษฎีที่เกี่ยวข้อง Complete 75 % (late: 51 วัน)** หมายเหตุ พักไว้สำหรับเพื่อในอณนาคตมีเรื่องที่จะต้องศึกษาเพิ่มเติม

**เตรียม Server Complete 50 % (late: 45 วัน)** หมายเหตุ จากที่ได้ Discuss กับที่ปรึกษาทำให้ต้องพักส่วนนี้ไปก่อนแล้วไปมุ่งเน้นกับการพัฒนาตัว Model Similarity

**เตรียม Data สำหรับการพัฒนา Complete 63 % (late: 30 วัน)** หมายเหตุเนื่องด้วยจำเป็นต้องมี data ที่เป็นแบบสอบถามการทำงานของนักศึกษาหลังจากเรียนจบ ซึ่งต้องใช้เวลาในการรอการตอบกลับของแบบสอบถาม

**พัฒนา Function NLP สำหรับจับกลุ่มวิชา Complete 100 %** หลังจากที่ผ่านการ Tuning model มาแล้ว และจากการแบ่งข้อมูลแบบ K-Fold Cross Validation โดยแบ่งเป็น 10 Fold โดยค่า Mean ของ Fold ทั้งหมดอยู่ที่ RMSE = 0.5646 และ MAE = 0.4309 ในกรณีของ Data ที่เป็นของนักศึกษาภาคปกติ และค่า Mean ของ Fold ทั้งหมดในกรณีของนักศึกษาภาคต่อเนื่อง RMSE = 0.7816 และค่า MAE = 0.6154 โดย Algorithm ที่จะใช้ในการสร้าง Model นี้คือ SVD (Singular Value Decomposition) หลังจากนั้นได้นำ code ที่ทดลองใน Google Collab ไปลงบน Backend Function ของ Demo Website ที่เตรียมไว้และทดสอบอีกครั้ง โดยค่า Mean ของ Fold ทั้งหมดอยู่ที่ RMSE = 0.5683 และ MAE = 0.4303 ในกรณีของ Data ที่เป็นของนักศึกษาภาคปกติ และค่า Mean ของ Fold ทั้งหมดในกรณีของนักศึกษาภาคต่อเนื่อง RMSE = 0.8070 และค่า MAE = 0.6388 แล้วได้ทำการ Implement ให้อยู่ในรูปแบบของการขอ Request prediction

อัพเดตรายงานบทที่ 1 **Complete 100 %**

****

*รูปที่ 1 บทที่ 1 บทนำ*

Text

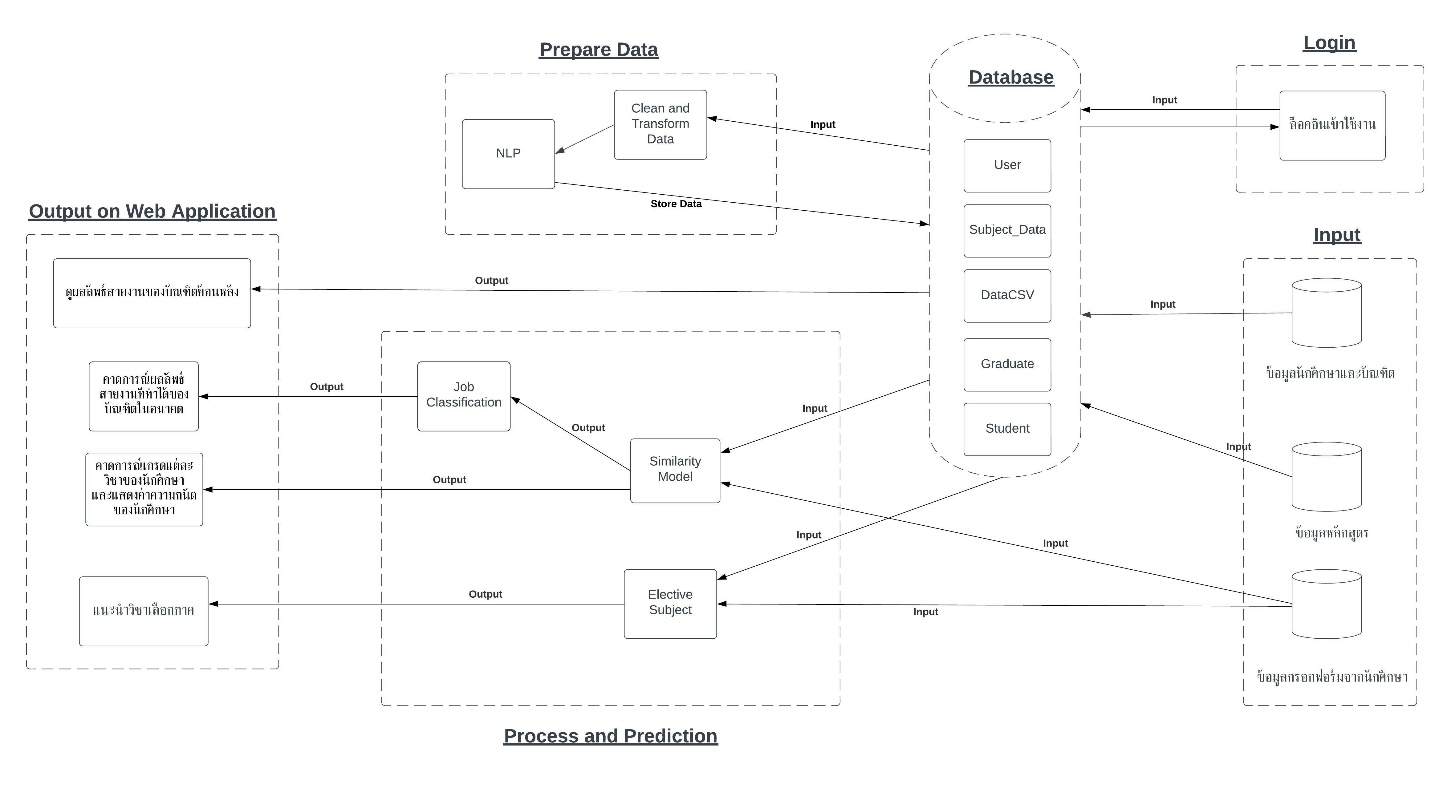
Description automatically generated

Timeline

Description automatically generated

*รูปที่ 2 บทที่ 1 บทนำ(ต่อ)*

อัพเดตส่วนของการออกแบบ



*รูปที่ 3 โครงสร้างการทำงานของระบบ*

โครงสร้างการทำงานของระบบได้อธิบายถึงการเชื่อมต่อระหว่างส่วนต่างๆของระบบ เริ่มตั้งแต่ส่วนของ Input ที่ทำหน้าที่รับข้อมูลเข้ามา แล้วเก็บไว้ในส่วนของ Database จากนั้นส่วนของ Prepare Data จะนำข้อมูลจากส่วนของ Database เมื่อทำเสร็จแล้ว จะทำการส่งกลับไปอัพเดตยัง Database ส่วนของ Process and Prediction จะนำข้อมูลที่ได้ไปเข้าอัลกอริทึมเพื่อ Process ผลลัพธ์ออกมาแสดงผลบน Web Application

จากรูป โครงสร้างการทำงานของระบบนั้นประกอบไปด้วยองค์ประกอบทั้งหมด 6 ส่วน ซึ่งมีรายละเอียด ดังนี้

1. Login

เป็นส่วนสำหรับไว้ให้กรรมการหลักสูตรได้ทำการล๊อคอินเข้าสู่ระบบเพื่อให้กรรมการหลักสูตรป้อนข้อมูลเกรดของนักศึกษาปัจจุบัน 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 และประมวลผลพร้อมทั้งตัดข้อมูลส่วนที่ไม่ได้นำมาใช้ออก

2. NLP หรือ Natural Language Processing

นำข้อมูลรหัสวิชามาเข้ากระบวนการ NLP เพื่อทำการหาค่า Similarity ของกลุ่มวิชาที่สามารถอยู่ในกลุ่มเดียวกันได้ โดยใช้บทคัดย่อของแต่ละวิชา เพื่อลดปัญหาการเปลี่ยนรหัสวิชาระหว่างหลักสูตร เมื่อทำเสร็จกระบวนการแล้ว จะนำข้อมูลที่ได้ กลับไปอัพเดตที่ Database

5. Process and Prediction

เป็นส่วนการประมวลผลหลักของระบบ ประกอบไปด้วย Process 3 ส่วนดังนี้

1. Similarity Model

มีหน้าที่นำข้อมูลผลการเรียนของนักศึกษามาทำให้สมบูรณ์ ด้วยกรณีของนักศึกษาแต่ละคนมีเกรดแต่ละวิชาที่ไม่เหมือนกัน บางคนมีเกรดของวิชานี้ แต่อีกคนไม่มีเกรดของวิชานี้ 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

## **ปัญหาที่เกิดขั้นและแนวทางการแก้ไข**

ปัญหาจากครั้งก่อน

1 NLP ในเรื่อง Similarity

จากการทดลองใช้งานจริงพบว่าการจับกลุ่มแบบนี้ทำให้บางวิชาที่ควรอยู่หมวดเดียวกันดันอยู่คนละหมวดเช่น Cal 1 และ Cal 2 นั้นจะอยู่กันคนละหมวดกันถ้าปรับการจัดหมวดหมู่อยู่ที่ มากกว่า 90% แต่ถ้าปรับอยู่ที่ 80% จะจัดหมวดหมู่กันได้ปัจจุบันได้แก้ไขปัญหาโดยการปรับให้ตัววิชาจับกลุ่มกันอยู่ที่ 90%

2 การเพิ่ม column subject type ให้ Dataframe นักศึกษา

เนื่องด้วยปัญหานี้เกิดจากที่ว่าในข้อมูลของนักศึกษานั้นมีวิชาที่ในข้อมูลของรายละเอียดวิชาไม่มีอยู่ทำให้เมื่อนำ column ของ subject types มาเพิ่มใน Dataframe ของที่จะ trian model นั้นเกิดเป็น Not Found Subject Type โดยปัจจุบันแก้ปัญหาโดยการใส้เป็นหมวดอื่นๆไปก่อน

ปัญหา ณ ปัจจุบัน

1. ข้อมูลในการสร้างโมเดลมีไม่เพียงพอ ทำให้การทำงานในส่วนของ job prediction ยังคงต้องรอข้อมูลจากอาจารย์ที่ปรึกษาก่อน

2. การเทรนข้อมูลมีการเทรนใหม่ทุกรอบ ทำให้ผลลัพธ์ออกมามีค่าที่ไม่นิ่ง

## **สิ่งที่จะดำเนินการต่อไป**

1. เขียนบทที่ 2 ของรายงานในส่วนของเอกสารที่เกี่ยวข้อง

2. ปรับแก้การเทรนใหม่ของโมเดล โดยหาทางปรับให้การทำงานทั้งหมดเร็วขึ้นและมีประสิทธิภาพมากขึ้น เช่น นำ Model ที่ train เสร็จแล้วนำไปเก็บไว้ในที่ ที่สามารถเรียกใช้งาน แก้ไข หรือ Maintains ตัวของ Model ได้ตลอดเวลา