# รายงานความก้าวหน้าวิชา CE Project

**ครั้งที่ 2**

**ระหว่างวันที่ 28 ส.ค. 65 ถึงวันที่ 09 ก.ย. 64**

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

## การดำเนินงานมีความก้าวหน้า 16 % (ใช้ค่า **% Complete** จาก MS Project)

มีความก้าวหน้าเพิ่มขึ้นจากรายงานความก้าวหน้า ครั้งก่อน 9 %

□ เร็วกว่าแผน 3 วัน □ ช้ากว่าแผน วัน

## รายละเอียดความก้าวหน้า

นัดประชุมกับที่ปรึกษาจำนวน 1 ครั้ง

ครั้งที่ 1 : หัวข้อการประชุม 2 หัวข้อ 1 Review Diagram แผนผังรวมของระบบ โดยมี Feedback : ต้องแก้ไขบางคำให้สื่อความหมายให้ชัดเจนมากขึ้น เช่น ข้อมูล file CSV ไปเป็น ข้อมูลเกรดของนักศึกษา แล้วปรับแก้ให้สอดคล้องกับ Usecase Diagram มากยิ่งขึ้น เช่นอาจต้องมีกล่องเป็น Background เพื่อครอบรวม usecase ว่าอยู่ใน Functional ไหนในระบบหลักบ้าง

2 ชี้แนะการดำเนินงานต่อไป : รวบรวมข้อมูลทั้งจากที่ ที่ปรึกษาจัดเตรียมให้ และ จัดเตรียมเองขึ้นไปทดลองใน Google Collab เพื่อปั่นเป็น Schema ตามที่ได้ Design ไว้ หลักจากนั้นจะเริ่มทำการ Data Processing ทั้ง Clean และ Transform เพื่อให้พร้อมต่อการใช้งาน แล้ว นำไปทดลองต่อตามที่ได้วางแผนไว้ใน Gantt Chart

หัวข้อการพัฒนาโครงงานตาม Gantt Chart

ศึกษาทฤษฎีที่เกี่ยวข้อง Complete 75 % (late: 9 วัน)

* ด้านเทคโนโลยี

หลังจากสืบค้นข้อมูลได้พบ library ตัวนึงของ python ที่หน้าสนใจคือ Surprise ซึ่งเป็น lib ที่อยู่ใน subset ของ Scikit learn อีกที ซึ่งถูกออกแบบมาเพื่อทำ Recommendation System โดยเฉพาะซึ่งสามารถเลือก Algorithm ได้หลากหลาย ซึ่งเหมาะกับตัวโครงงานเป็นอย่างมากเลยเลือกที่จะใช้ lib ตัวนี้เป็น lib หลัก

* ด้านงานวิจัยที่เกี่ยวข้อง

ได้ทำการ Review แล้วทั้ง 2 งานวิจัยสรุปได้ว่า

1 สุเมธ ดาราพิสุท นำเสนองานวิจัยเรื่อง การสร้างรายการเพลงโดยใช้การกรองร่วมแบบเซสชั่นที่เพิ่มขึ้นด้วยกลไกการลืมและการวิเคราะห์สถิติเชิงมุม โดยใช้ 2 วิธีร่วมกัน 1 การสร้างรายการเพลงจะพิจารณาการฟังเพลงในเซสชั่นปัจจุบันที่คล้ายกับเซสชั่นในอดีตของผู้ฟัง 2 สร้างรายการเพลงแนะนำโดยพิจารณาช่วงเวลา เฉพาะในการฟังเพลงซึ่งแตกต่างจากช่วงเวลาอื่นอย่างมีนัยสำคัญทางสถิติในรอบวันของผู้ฟังโดยใช้ การวิเคราะห์สถิติเชิงมุม

และวัดประสิทธิภาพโดย : ประสิทธิภาพ HitRatio และ Precision จากการทดลองพบว่าการใช้ 2 วิธีแยกกันนั้นได้ผลลัลพท์ที่น้อยกว่านำมาใช้ร่วมกัน 0.18-0.22 %

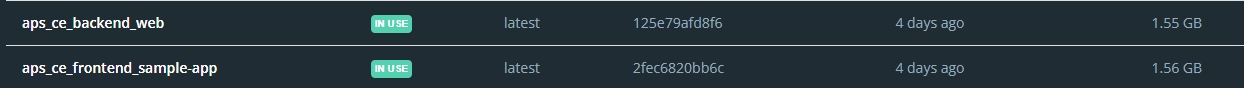
2 นิภาภรณ์ พันธ์นาม นำเสนองานวิจัย ระบบแนะนำสินค้าอาหารโดยใช้ระบบแนะนำแบบผสมผสาน

ใช้เทคนิค Content based filtering แบบหลักการ Cosine และสร้างแบบจำลองโดยใช้ lib Surprise ซึ่งมีอัลกอริทึม SVD, NMF, Baseline และ KNN และวัดประสิทธิภาพโดย RMSE, MAE จากการทดลองพบว่า 1 เทคนิคการกรองแบบอิงเนื้อหานำวิธีการ TF-IDF เข้ามาช่วยในการทำ Vectorization ส่วนใหญ่ค่าความเหมือนออกมาค่อนข้างที่จะตำเนื่องมาจากข้อมูลที่น้อยเกินไป 2 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม ผ่าน library Surprise ของ Scikitlearn ซึ่งโมเดลที่มีผลคะแนนโดยรวมดีที่สุดคืออัลกอริทึมของ SVD ซึ่ง ได้ค่ำ RMSE 1.2528 และ MAE 0.9376 และ 3 ระบบแนะนำแบบผสมผสาน โดยผลลัพท์นั้นจะไม่ชัดเจนเนื่องจากวิธีนี้ไม่ได้ เนื่องจากกรณีนี้ได้มีการทำนายค่า Rating ซึ่งวิธีการของระบบแนะนำแบบผสมผสานนั้น ได้มีนำเทคนิคการกรองแบบอิงเนื้อหา

ที่ไม่ได้มีการทำนายค่าอะไรมารวมในการทำงานของแบบจำลองด้วย ซึ่งถ้าต้องการวัดผลลัพท์สามารถอ้างอิงจากค่า RMSE, MAE ได้

เตรียม Docker Complete 100% (early: 3 วัน)

โดยจากครั้งก่อนที่มีการรวมทุกอย่างไว้ใน Image เดียวนั้นได้เปลี่ยนมาเป็นการแยกส่วนของ Image ออกมาเป็นส่วนของ Backend ที่ใช้ Framework ของ Django และ มี Image ของ Database Server เป็น SQL โดยใช้ MariaDBและ Image ของ Frontend ที่ใช้เป็น React

ซึ่งหลังจากวางโครงสร้างและสรุปการทำงานกันจะได้ว่าจะนำ Image ของ Backend และ Image ของ Database Compose ขึ้นไปเป็น Container พร้อมกัน โดยจะเปิด Port สำหรับ Development อยู่ที่ 8000 สำหรับ Backend และ 3306 สำหรับ Database Server และในส่วนของ Image Frontemd นั้นได้ทำเป็น Compose เดี่ยวขึ้นไปและทำการเปิด Port สำหรับ Development อยู่ที่ 3000

*รูปที่ 1 Docker Image ที่ใช้งานหลักสำหรับ Wep Application*

Graphical user interface, application

Description automatically generated*รูปที่ 2 Docker Compose Container Frontend, Backend และ Database Server*

เตรียม Server Complete 50 % (late: 1 วัน)

จากการเปลี่ยนแผนใหม่ทำให้ยังไม่มีอะไรคืบหน้านอกจากทดลองนำ Docker Compose ไปสร้าง Container บน Server

เตรียม Data สำหรับการพัฒนา Complete 72 %

หลังจากได้ Data มาจากที่ปรึกษาและที่หามาเองได้เอา Data มา Clean เช่นใน Data เกรดและวิชาของ นักศึกษา column year นั้นไม่ได้เป็นประโยชน์จึง Drop ทิ้งไปและได้ Transform โดยใน Column Grade ได้เปลี่ยนจากระบบ Char (A, B+, B, …, F) ไปเป็น Int(4, 3.5, 3, …, 0) และได้ให้เปลี่ยนเกรด S ให้มีค่าเท่ากับ 4 และ U เท่ากับ 0 หลังจากนั้นได้ใช้ pandasql ในการ query จากตารางหลักออกมาเป็นตารางย่อยที่จะนำไปใช้ประโยชน์ต่อไป

A picture containing table

Description automatically generated

*รูปที่ 3 Transform Data Drop column ที่ไม่ใช้ Computer Engineer ภาคปกติและต่อเนื่อง*

Graphical user interface, text

Description automatically generatedA picture containing text

Description automatically generatedA picture containing table

Description automatically generatedText

Description automatically generated *รูปที่ 4 Transform Grade ให้อยู่ในรูปตัวเลข* *รูปที่ 5 Query ข้อมูลนักศึกษา*

*รูปที่ 6 Query ข้อมูลรายวิชาต่างๆและเพิ่ม Delete Flag*

*รูปที่ 7 load ข้อมูลมาเป็น csv*

## ปัญหาที่เกิดขึ้นและแนวทางการแก้ไข

ปัญหาจากครั้งก่อน

1 Task และ Workload

ปัจจุบันทั้งงานและ workload นั้นเป็นไปตามที่เหมาะสมเรียบร้อยแล้ว

2 Data Processing

ปัจจุบันได้มีการวางแผนและแตก task ที่ต้องทำออกมาอย่างมีระบบและขั้นตอนแล้ว

ปัญหา ณ ปัจจุบัน

1 Docker Image Design

เนื่องจากตอนแรกได้ Design ให้ Docker Image นั้นรวบร่วมนำสิ่งที่ต้องใช้งานไว้ใน Image เดียวกันทำให้เกิดความสับสนในการวาง port การติดต่อและ development code ได้ยาก จึงได้แก้ปัญหาโดยการแยกนำ Backend และ Database และ Frontend ออกมาเป็นอย่างละ Image ซึ่งได้ทำการ Compose รวม Backend และ Database สร้าง Compose Container ขึ้นมาเพื่อให้ Django setup SQL Database Server ได้ง่ายยิ่งขึ้น และแยก Frontend เป็น Compose Container เนื่องจากการ Run เป็น Compose up นั้นจะทำให้ผู้จัดทำ ทำงานได้ถนัดกว่าเป็น Image ซึ่งจะยังไม่ได้ทดลองติดต่อกับ Database Server ผ่านตัว Backend ทำให้ยังไม่ทราบว่าจะประสบกับปัญหาใดบ้าง

2 Data Transform

เนื่องจากตัว Algorithm ส่วนมากใน Recommendation ของ Surprise นั้นจำเป็นต้องใช้ข้อมูล Rating ที่เป็นตัวเลขจึงต้องแก้ไข Grade ของนักศึกษาที่เป็นรูปแบบ Char (A, B+, B, …, F) ไปเป็น Int(4, 3.5, 3, …, 0) ซึ่งปัญหาคือไม่สามารถทราบได้ว่าจะนำเกรด S, U ไปเทียบกับเลขใด โดยปัจจุบันแก้ปัญหาโดยการ เปลี่ยนเกรด S ให้มีค่าเท่ากับ 4 และ U เท่ากับ 0

## สิ่งที่จะดำเนินการต่อไป

* ทำการศึกษาทฤษฎีที่เกี่ยวข้องให้ครบเพื่อนำไปประกอบกับรายงานหลัก
* ทำการ Map, Transform, Clean ข้อมูลให้ครบตาม Design ที่ได้ทำไว้ ผ่าน Google Collab
* ทดลองนำข้อมูลที่ได้เตรียมมาใช้กับ library Surprise ของ Scikit learn เพื่อเลือก algorithm ที่ดีที่สุด