

รายงานความก้าวหน้าวิชา CE Project 2

ครั้งที่ 4

ระหว่างวันที่ 23 ก.พ. 66 ถึงวันที่ 24 มี.ค. 66

1. ชื่อโครงการ (อังกฤษ) Curriculum Output Prediction From Student Academic Data
2. การดำเนินงานมีความก้าวหน้า 78 % (ใช้ค่า **% Complete** จาก MS Project)

มีความก้าวหน้าเพิ่มขึ้นจากรายงานความก้าวหน้า ครั้งก่อน 13 %

☐ เร็วกว่าแผน วัน ☒ ช้ากว่าแผน 27 วัน

3. รายละเอียดความก้าวหน้า

นัดประชุมกับที่ปรึกษาจำนวน 2 ครั้ง

ครั้งที่ 1 : การประชุมในครั้งที่ 1 ได้มีหัวข้อการประชุมดังนี้ 1) พูดคุยถึง requirement progress 2) แนวทางในการพัฒนาต่อ

- 1) พูดคุยถึง requirement progress ทางผู้จัดทำได้อธิบายแนวทางการพัฒนาในปัจจุบันให้ทางที่ปรึกษาโดยมีรายละเอียดดังนี้ โดยปัจจุบันงานจะเน้นการพัฒนาในส่วนของหน้าการใช้งานของ web application หรือ frontend โดยทางการพัฒนาในส่วนการทำงานหลังบ้านหรือ backend นั้นจะเป็นการปรับแต่ง algorithm ของการทำงานให้ง่ายต่อการ scalability มากขึ้นและขนานไปกับตัว design ของ frontend
- 2) แนวทางการพัฒนาต่อจากที่ได้พูดคุยกับทางที่ปรึกษานั้น ทางที่ปรึกษาได้แนะนำให้ทางผู้จัดทำนั้นดังนี้โดยได้มีการพูดคุยกับที่ปรึกษาเรื่องการปรับเปลี่ยน algorithm การทำงานของการ prediction ในส่วนของการทำนายอาชีพของนักศึกษาโดยปรับเปลี่ยนการ process input data และขั้นตอนการ train และ generate model โดยเพื่อปรับ accuracy performance ของตัวระบบให้สูงขึ้น

ครั้งที่ 2 : การประชุมในครั้งที่ 2 ได้มีหัวข้อการประชุมดังนี้ 1) พูดคุยถึง requirement progress 2) แนวทางในการพัฒนาต่อ

- 1) พูดคุยถึง requirement progress ทางผู้จัดทำได้อธิบายแนวทางการพัฒนาในปัจจุบันให้ทางที่ปรึกษาโดยมีรายละเอียดดังนี้ โดยปัจจุบันงานจะเน้นการพัฒนาในส่วนของหน้าการใช้งานของ web application หรือ frontend โดยทางการพัฒนาในส่วนการทำงานหลังบ้านหรือ backend นั้นจะทำการ pause process ไประยะหนึ่งเพื่อมุ่งเน้นไปที่การ accuracy performance tuning ของตัว algorithm process ที่ใช้ โดยได้มีการแจ้งผลการทดลองให้กับที่ปรึกษาไปว่า จากผลลัพธ์การทดลองก่อนเปลี่ยน algorithm นั้นวัดผลค่า accuracy ของ model อยู่ที่ 0.2 หรือ 20% โดยหลังจากทำการพัฒนา algorithm ในการ process ใหม่อีกครั้งแล้ววัดผลค่า accuracy ของ model จะอยู่ที่ 0.4 หรือ 40%

- 2) แนวทางในการพัฒนาต่อจากผลลัพธ์การพัฒนาที่ได้ออกมานั้นยังไม่เป็นที่น่าพอใจจากที่คาดการณ์เอาไว้โดย accuracy ที่ทางที่ปรึกษาได้คาดการณ์ไว้นั้นอยู่ที่ 0.7-0.9 accuracy หรือ 70-90% โดยทางที่ปรึกษาได้แนะนำให้ปรับเปลี่ยนตัว algorithm ในการ process เพิ่ม ส่วนการพัฒนาในส่วนของ front end นั้นยังคงเป็นไปตามแผน กำหนดการเดิม โดยทางที่ปรึกษาได้มี concern อยู่คือให้ปรับ usage การใช้งานให้เป็นเชิง technical น้อยลง

หัวข้อการพัฒนาโครงการตาม Gantt Chart

ศึกษาทฤษฎีที่เกี่ยวข้อง Complete 100 % หมายเหตุ ในส่วนการพัฒนาในปัจจุบันนั้นไม่ทฤษฎีที่ต้องศึกษาเพิ่มเติม

เตรียม Data สำหรับการพัฒนา Complete 100 % หมายเหตุ เนื่องด้วยในช่วง progress ที่ผ่านมาทางที่ปรึกษาได้ทำการนำข้อมูลที่ทางผู้จัดทำได้รวบรวมมาจากการทำแบบสอบถามการมีงานทำของบัณฑิตมาทำการมาเข้ารหัสข้อมูลรหัส นักศึกษาหลังจากนั้นทางผู้จัดทำได้นำข้อมูลในส่วนนั้นมาเพิ่มลงใน column ของ student_data

พัฒนา Job Classification Model Complete 95 % หมายเหตุเนื่องด้วยจากข้อสรุปที่ได้ประชุมกับทางที่ปรึกษานั้น จำเป็นต้องปรับเปลี่ยน algorithm ในการ process data ใหม่ โดยแนวทางในการปรับการ accuracy performance นั้นจะมีด้วยกัน 2 ครั้ง โดยครั้งแรกจะทำการปรับโดยใช้เพียงแค่ fact data เท่านั้นในการ train model โดยจะทำการ filter main data set depend on กับข้อมูลของของผู้ใช้งานที่ร้องขอการทำนาย รวมกับการใช้ algorithm ในการ prediction เป็น Randomforest Classifier โดยในครั้งที่ 2 นั้นจะทำการปรับลดจำนวนของ class ที่ต้อง prediction ให้ลดลง จาก 18 เป็น 10 โดยตัวอย่างของการ fact data ที่ผ่านการ filter มานั้นจะมีลักษณะตามรูปที่ 1

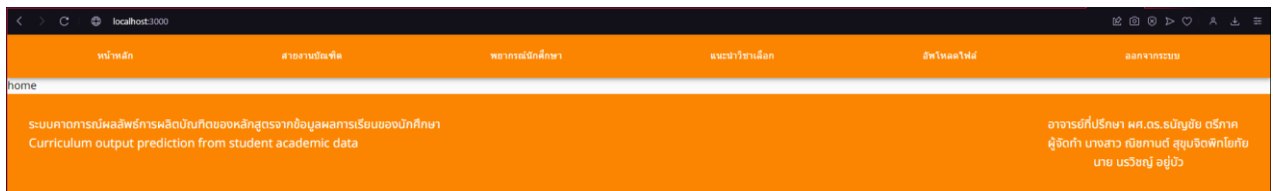
พัฒนา Django Web App Complete 65% (remaining 39.64 Hr) หลังจากที่ได้ทำการคุยแนวทางการพัฒนาต่อกับที่ปรึกษาในทั้ง 2 การประชุมนั้นในส่วนของ backend จะยังไม่มี progress เพิ่มเติมนอกจากการปรับ logic เพื่อให้สอดคล้องกับ front end

	student_id	1006028	1006030	1006031	1076001	1076002	1076003	1076004	1076005	1076006	...	1076008	1076009	1076010	1076011	1076012	1076027	1076032	1076253	1076263	job
15	bb4e051171993c63c9e0717260052615	4.0	3.0	2.5	3.5	3.0	2.5	3.0	1.5	3.0	...	2.5	3.0	2.5	3.0	1.5	2.5	2.5	2.5	3.5	Software Engineer
20	3ac0de4c4710ac58f180405f45da8205	4.0	2.5	3.0	3.5	3.0	2.5	3.5	3.0	3.0	...	4.0	3.0	2.5	3.0	3.0	3.5	3.0	3.5	4.0	Software QA
23	d7f1edbe586f9607c1a18004e60310f	4.0	3.0	2.5	3.5	2.0	2.0	2.0	2.0	2.5	...	3.0	3.5	2.5	3.5	4.0	2.0	3.5	2.5	3.5	Software Developer
30	f1897be137c1342811f9007fbf100416	4.0	2.0	2.0	3.5	3.5	1.5	2.5	3.5	3.0	...	3.5	3.5	3.5	3.5	3.5	2.0	2.5	3.5	3.5	Mobile Developer
32	ea3069733251752fa24e430e3d73ec94	4.0	1.5	2.5	3.5	2.5	1.5	2.0	3.5	2.0	...	2.0	2.5	2.5	2.5	2.5	2.0	1.0	2.0	2.5	Network Administrator
34	3b387161d63cc3a3e4f643b339c2072a0	4.0	3.0	3.5	4.0	4.0	2.5	4.0	3.5	3.5	...	3.0	3.0	3.0	3.5	3.5	3.0	2.5	3.5	3.0	Software Engineer
39	15eef5c71ce09039ae9c104849bd3e47	4.0	2.0	3.0	3.5	2.5	2.5	2.5	3.0	3.0	...	2.5	3.5	4.0	3.0	2.0	2.5	2.5	3.5	3.5	Software Developer
44	6f3da1068e00358c8ec72513300dba8f	4.0	2.0	2.5	3.0	1.5	1.5	1.5	1.5	2.0	...	2.0	3.0	2.5	2.0	1.5	1.5	2.0	2.5	2.0	UX UI Designer
47	c7605751fa1a39f5d7e35159030b5211	4.0	2.0	2.0	4.0	3.0	4.0	3.5	4.0	4.0	...	2.5	4.0	3.0	4.0	3.0	4.0	3.5	2.5	3.0	Firmware Developer
49	0f6119e2880d31fb88ceddb6554573	4.0	2.5	4.0	2.5	1.5	2.0	2.5	1.0	2.0	...	3.5	3.0	3.0	3.5	3.0	2.5	2.0	3.0	3.0	Data Engineer
51	7d5203d2d37502c500d662e47e1644	4.0	2.0	3.5	4.0	3.5	3.5	4.0	3.5	4.0	...	4.0	4.0	3.0	4.0	3.0	3.5	4.0	3.0	3.5	Software Engineer
53	1e28bfce0b9e3c2028f79cb33365db1	4.0	1.5	2.0	3.5	2.0	2.5	3.0	2.5	3.5	...	3.5	3.0	2.5	3.0	2.5	2.5	3.5	3.0	3.5	IT Infrastructure
58	78fee4606c034f4785e9e1a827e4fca	4.0	3.5	4.0	4.0	4.0	3.5	4.0	4.0	4.0	...	3.5	4.0	4.0	4.0	4.0	4.0	2.5	3.5	3.5	Software Engineer
59	0a079386e7934625831ece1087ae915f	4.0	2.5	2.0	3.5	3.5	2.5	3.5	3.0	3.5	...	3.0	3.0	3.0	3.0	3.5	2.5	2.5	3.0	3.0	Mobile Developer
61	9a584c9433b0dce2524bdb54064c925	4.0	3.0	3.0	3.5	3.0	2.5	3.5	2.5	4.0	...	2.5	3.5	3.0	3.5	3.5	2.0	2.5	3.5	3.5	Software Engineer
62	2b25ba620f6313a62a0727fdae38d13e	4.0	2.5	1.5	4.0	3.0	2.5	2.5	3.0	2.5	...	3.0	2.5	2.5	3.5	3.5	3.5	2.0	2.5	3.0	Software Developer
65	b7da48f490db705e21b25a09259df92	4.0	3.0	2.5	3.5	2.0	2.0	3.0	2.0	2.0	...	3.5	3.0	4.0	2.5	2.0	2.5	3.0	3.0	3.5	Frontend Developer
72	a00b4e2d58030dede1e0020f47521ae6	4.0	2.0	3.5	3.5	2.5	2.5	2.5	2.5	3.5	...	3.0	3.5	3.5	2.5	3.0	2.5	2.5	2.5	2.5	UX UI Designer
73	cfc97a31724ea1f87e31b5992a3e527c	4.0	1.5	3.5	3.5	2.0	2.0	3.0	2.0	2.5	...	3.0	3.0	3.5	3.0	2.5	3.0	3.0	3.0	3.5	UX UI Designer
76	65365996c164fe1983a09e6ef3687e	4.0	2.0	2.0	3.0	1.0	2.0	2.0	2.5	3.5	...	3.0	3.5	2.5	3.0	2.0	2.0	1.5	3.5	3.5	UX UI Designer
78	44ae02b1e659304dc84bbcb9ab2dced5	4.0	3.0	3.0	3.5	3.5	2.5	3.5	3.0	3.5	...	2.5	3.0	2.5	3.0	4.0	2.5	2.5	2.5	3.0	Software Developer
79	0faeb665c528f3c41f327a57a8575c	4.0	2.5	3.0	3.5	3.0	3.0	3.5	3.5	3.0	...	3.0	3.5	3.0	3.5	2.5	3.0	3.5	2.0	3.5	Software Tester
82	839f9b47419d9b8f442514d1da429e34	4.0	1.5	3.0	3.5	3.0	1.5	3.0	2.5	2.5	...	3.5	3.5	3.5	3.0	2.5	3.0	2.0	4.0	4.0	Software Engineer
85	a208cc29835c203e85bd2e42897b654f	4.0	3.5	3.0	3.0	3.5	2.5	3.0	3.0	3.0	...	3.5	4.0	3.0	3.0	3.5	2.5	2.5	3.0	3.5	Data Engineer

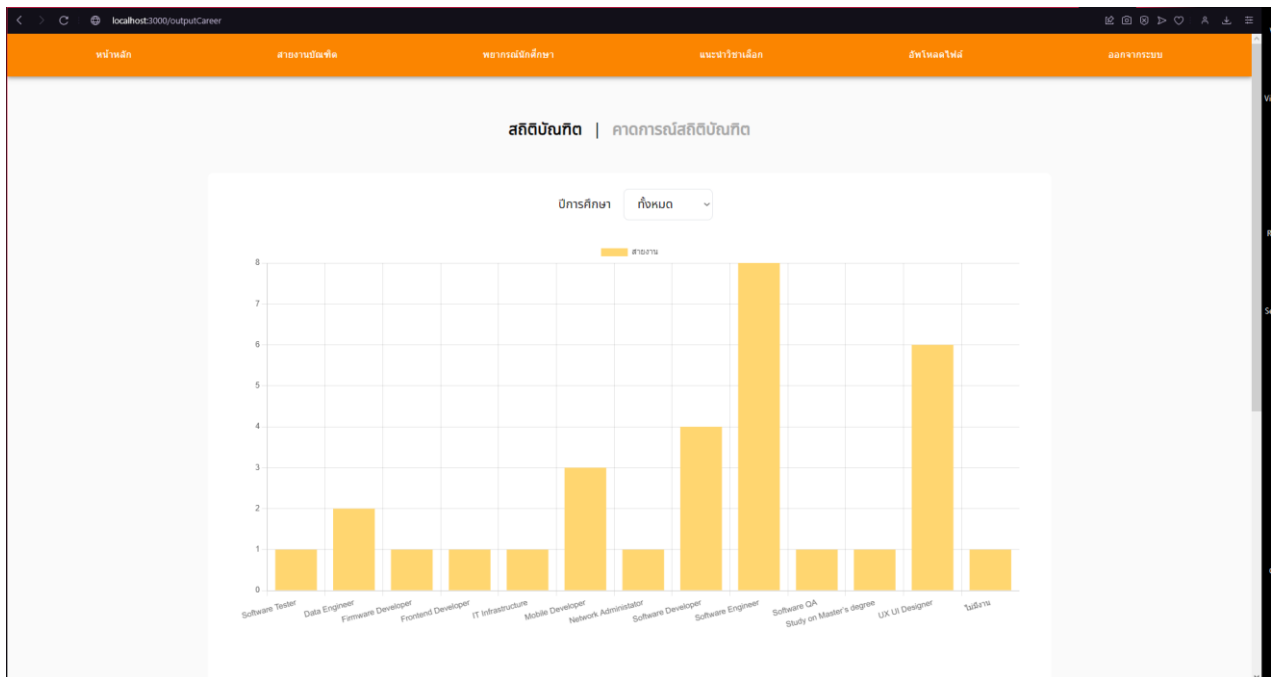
24 rows x 21 columns

รูปที่ 1 ตัวอย่าง fact data ที่ผ่านการ process แบบใหม่

พัฒนา Frontend Complete 75% (remaining 10 Hr) การพัฒนาของ frontend นั้นจะมีความก้าวหน้าตามรูปภาพ



รูปที่ 2 หน้าหลักของ web application



รูปที่ 3 หน้าแสดงจำนวนอาชีพทั้งหมด

รูปที่ 4 หน้าใช้งานระบบการทำนายอาชีพ

รูปที่ 5 หน้าใช้งานระบบการแนะนำรายวิชา

รูปที่ 6 หน้าใช้งาน upload file

รูปที่ 7 หน้าใช้งานระบบ login

4. ปัญหาที่เกิดขึ้นและแนวทางการแก้ไข

ปัญหาจากครั้งก่อน

1. ปัญหาการ prediction ที่ไม่สมบูรณ์ จากครั้งก่อนที่ทางผู้จัดทำได้ investigate issue แล้วได้ตั้งสมมติฐานว่าข้อผิดพลาดอาจเกิดจากการทำงาน 2 ส่วน ส่วนแรกคือการ process data ที่จะใช้ในการ train model ส่วนที่สองคือส่วนในการนำข้อมูลเข้า prediction ซึ่งหลังจากทำการ research เพิ่มเติมพบว่า use case ในการใช้งานการร้องขอการ prediction ของระบบที่ทางผู้จัดทำนั้นได้ออกแบบมานั้นไม่ได้สอดคล้องกับการใช้งานจริงของ model svd surprise scikit เนื่องจากตัว svd สามารถ prediction ได้แค่เพียง user ที่อยู่ใน data set ที่ใช้สำหรับ train model เท่านั้นแต่ usecase ของระบบนั้นเป็นการถามหาคำทำนายของ user ที่ไม่อยู่ในระบบจึงได้ค่า prediction ที่เหมือนเดิมเสมอเมื่อมี pattern ของ behavior ที่คล้ายกันแม้ rating จะต่างกัน โดยทางผู้จัดทำได้ทำการแก้ปัญหาด้วยการ redesign ตัว process นี้ใหม่ด้วยการ trade off runtime performance กับ accuracy โดยยอมสร้าง dataset ใหม่แล้ว train model ใหม่ทุกครั้งที่มีการร้องขอการทำนาย ผลลัพธ์ accuracy ออกมาเป็นที่น่าพอใจโดยอยู่ในช่วง 0.49 – 0.65 rmse และ run time per request อยู่ที่ 15 วินาทีโดยเฉลี่ยบน local host ซึ่งถือว่าเป็นอยู่ในเกณฑ์ที่พอรับได้โดยปัญหานี้ก็จบด้วยดี

ปัญหา ณ ปัจจุบัน

1. ปัญหา accuracy performance ของ model job classification เนื่องจากข้อสรุปการประชุมจากที่ทางที่ปรึกษาต้องการให้ทำการปรับค่า accuracy ให้อยู่ที่ 0.7-0.9 หรือ 70-90% โดยทางผู้จัดทำได้ตั้งสมมติฐานว่าข้อผิดพลาดอาจเกิดจากที่ตัว model นั้นไม่สอดคล้องกับ data ที่ process มา, data ที่ใช้นั้นเป็น data ที่ผ่านการ prediction มาก่อน ไม่ใช่ fact data และ จำนวนของ data ที่มีใช้ในการ train นั้นน้อยเกินไป โดยทางผู้จัดทำได้แก้ปัญหาในรอบแรกด้วยการเลือกใช้ model ใหม่เป็น Randomforest Classifier แล้วทำการ pre process data ด้วยการเลือกใช้แค่ fact data โดยจัดการกับ null value ด้วยการ filter main dataset depend on data ของ user ที่ร้องขอคำทำนายโดยผลลัพธ์ออกมาดีขึ้น จาก 0.2 ไปเป็น 0.4 แต่ยังไม่เป็นที่น่าพอใจมากพอ ทางผู้จัดทำเลยได้ตั้งสมมติฐานเพิ่มเติมว่าอาจเกิดจากการที่ dataset นั้นมี class ที่แตกต่างกันมากเกินไปร่วมกับ feature ที่มากด้วยอยู่แล้ว โดยแนวทางในการแก้ไขปัญหานี้ว่าจะทำการปรับลด class ที่จะทำการ prediction ลงจาก 18 เหลือ 10 class

5. สิ่งที่จะดำเนินการต่อไป

1. ทำการพัฒนาโครงงานต่อตามแนวทางที่ได้ปรึกษากับที่ปรึกษา
2. จัดการปัญหาการ accuracy performance ของ model job classification ด้วยการปรับลด class ที่จะทำการ prediction ลงจาก 18 เหลือ 10 class
3. พัฒนา ในส่วนของ frontend