



โครงการทางวิทยาการคอมพิวเตอร์ระดับปริญญาตรี

เรื่อง

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยี
ปัญญาประดิษฐ์

Computer Internship Recommendation System With Artificial
Intelligence Technology

โดย

นายทินกฤต สิงห์แก้ว

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน
ปีการศึกษา 2565



ใบรับรองโครงงานวิทยาการคอมพิวเตอร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน^๑
วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)

เรื่อง ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์
Computer Internship Recommendation System With Artificial Intelligence Technology
โดย นายพินกฤต สิงห์แก้ว

คณะกรรมการพิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษา..... วันที่...../...../.....
(ผู้ช่วยศาสตราจารย์ ดร.มนูช เกตุย)

อาจารย์ที่ปรึกษาร่วม..... วันที่...../...../.....
(อาจารย์วรวิทย์ ผันคำข่าย)

อาจารย์ผู้รับผิดชอบวิชา..... วันที่...../...../.....
(อาจารย์ปกรณ์ สุนทรเมธ)

ประธานหลักสูตร..... วันที่...../...../.....
(อาจารย์วรวิทย์ ผันคำข่าย)

โครงการทางวิทยาการคอมพิวเตอร์ระดับปริญญาตรี

เรื่อง

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์

Computer Internship Recommendation System With Artificial Intelligence

Technology

โดย

นายทินกฤต สิงห์แก้ว

คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน

เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรบัณฑิต สาขาวิทยาศาสตร์

พ.ศ. 2565

บทคัดย่อ

ชื่อโครงการ	: ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์
ผู้ศึกษา	: นายทินกฤต สิงห์แก้ว
อาจารย์ที่ปรึกษา	: ผู้ช่วยศาสตราจารย์ ดร. นงนุช เกตุย
อาจารย์ที่ปรึกษาร่วม	: อาจารย์วรรวิทย์ ผั้นคำข้าย
สาขาวิชา	: วิทยาศาสตร์
หลักสูตร	: วิทยาการคอมพิวเตอร์
ปีการศึกษา	: 2565

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ (Computer Internship Recommendation System With Artificial Intelligence Technology) เป็นระบบที่ช่วยแนะนำสถานประกอบการสำหรับการฝึกงานของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ผ่านในรูปแบบของเว็บแอปพลิเคชันที่พัฒนาจากเว็บเฟรมเวิร์ค Next.js (Web Framework) สำหรับพัฒนาเว็บแอปพลิเคชัน โดยให้นักศึกษาระบุรายละเอียดความสนใจตามรูปแบบคร่าวๆ หรือรูปแบบของงานเพื่อนำมาวิเคราะห์ความคล้ายคลึงกับข้อมูลสถานประกอบการที่มีอยู่ในฐานข้อมูลที่ผ่านกระบวนการวิเคราะห์ความคล้ายคลึงของตัวที่คนหาจากเทคโนโลยีประมวลผลภาษาธรรมชาติ และจัดตามกลุ่มด้วยเทคโนโลยีการจัดกลุ่ม (K-Means) อยู่ในกลุ่มของการให้คอมพิวเตอร์เรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) ซึ่งเป็นส่วนหนึ่งของเทคโนโลยีปัญญาประดิษฐ์ โดยผลการทดสอบพบว่า นักศึกษาได้ใช้ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ (Computer Internship Recommendation System With Artificial Intelligence Technology) และได้มีประสิทธิภาพที่ค่าเฉลี่ยค่าความพึงพอใจอยู่ที่ 4.01 ซึ่งอยู่ในระดับดี

กิจกรรมประจำ

โครงการทางวิทยาการคอมพิวเตอร์ “ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ด้วยเทคโนโลยีปัญญาประดิษฐ์” (Computer Internship Recommendation System With Artificial Intelligence Technology) เพื่อการสำเร็จการศึกษาระดับปริญญาตรี สามารถดำเนินการจนประสบความสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาและคำแนะนำจากคณาจารย์หลาย ๆ ท่านในหลักสูตรวิทยาการคอมพิวเตอร์ ที่ได้กรุณาให้ความชี้แนะแนวทาง ขอคิด ขอแนะนำสู่ความสำเร็จและช่วยแก้ไขปัญหาอุปสรรคต่าง ๆ รวมทั้งรูปเล่มให้สำเร็จไปได้ด้วยดี

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.นงนุช เกตุย และอาจารย์วรวิทย์ ผึ้นคำข่าย อาจารย์ที่ปรึกษา และอาจารย์ประจำวิชาทุกท่าน ผู้ชี้ช่องกรุณาให้ความชี้แนะแนวทางการสร้างผลงานสู่ความสำเร็จ และช่วยแก้ไขปัญหาอุปสรรคต่าง ๆ รวมทั้งตรวจสอบแก้ไขรูปเล่มจนเสร็จสมบูรณ์ ขอขอบคุณบิดา มารดา ผู้มีพระคุณทุกท่าน เพื่อนนักศึกษา และบุคคลที่เกี่ยวข้องที่ยังไม่ได้กล่าวถึง ที่ได้ช่วยออกแบบคิดเห็น ได้ให้ข้อแนะนำ และอำนวยความสะดวกในด้านต่าง ๆ ในการทำโครงการครั้งนี้ได้ ณ ที่นี่

สุดท้ายนี้ผู้ศึกษาหวังว่าโครงการฉบับนี้จะเป็นประโยชน์สำหรับมหาวิทยาลัย และนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน และผู้ที่สนใจที่จะศึกษาต่อไป

สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ข
สารบัญ	ค
สารบัญตาราง	ง
สารบัญภาพ	ฉ
บทที่ 1 บทนำ	4
1.1 ความเป็นมาและความสำคัญของปัญหา	4
1.2 วัตถุประสงค์	4
1.3 ขอบเขตของโครงการ	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ	5
1.5 อุปกรณ์เครื่องมือที่ใช้ในโครงการ	5
1.6 นิยามศัพท์ที่ใช้ในโครงการ	6
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	7
2.1 ทฤษฎีและความรู้ที่เกี่ยวข้อง	7
2.2 งานวิจัยที่เกี่ยวข้อง	19
บทที่ 3 ขั้นตอนการดำเนินงาน	21
3.1 การเตรียมและวิเคราะห์ข้อมูล	21
3.2 การทำงานของระบบ	23
3.3 การวิเคราะห์และออกแบบระบบ	23
3.4 การออกแบบฐานข้อมูล	30
3.5 การออกแบบหน้าจอ	32
บทที่ 4 ผลการดำเนินงาน	36
4.1 การวิเคราะห์และการทำ Word segmentation	36
4.2 ขั้นตอนการใช้งานสำหรับผู้ใช้งาน	49
4.3 การวัดค่าความคล้ายคลึง	52
บทที่ 5 สรุปผลการดำเนินงาน	54
5.1 สรุปผลการดำเนินงาน	54
5.2 สรุปปัญหาที่เกิดระหว่างการดำเนินงาน	54
5.3 แนวทางพัฒนาระบบในอนาคต	55
5.4 แบบประเมินความพึงพอใจของผู้ใช้	55
เอกสารอ้างอิง	59
ภาคผนวก ก คู่มือการติดตั้งระบบ	62
1. การจัดกลุ่มข้อมูลด้วยเทคนิคทางปัญญาประดิษฐ์เคมีน (K-Means)	63
2. การสร้างคลัสเตอร์ (Cluster) ของฐานข้อมูลมองโภตีปี (MongoDB) บนเว็บไซต์	66
3. การสร้างฐานข้อมูลในโปรแกรมมองโภตีปี (MongoDB)	68
4. การนำข้อมูลเข้าสู่ฐานข้อมูลมองโภตีปี (MongoDB)	71

	๙
5. การใช้งาน Cloud computing ของ Amazon Web Services	74
6. การติดตั้งและใช้งาน Web API สำหรับคำนวณค่าความคล้ายคลึง (Cosine similarity)	76
7. การติดตั้งและใช้งานเว็บแอปพลิเคชัน (Web application)	77
ภาคผนวก ๖ คู่มือการใช้งาน	80
1. ผู้ดูแลระบบ	81
2. ผู้ใช้งาน	84
ประวัติผู้ศึกษา	56

สารบัญตาราง

ตารางที่	หน้า
1. ตัวอย่างการคำนวณค่า Term Frequency ที่จำนวนคำทั้งหมดเท่ากับ 7	9
2. ตัวอย่างการคำนวณค่า Inverse Document Frequency ที่จำนวนเอกสารเท่ากับ 10	10
3. ตัวอย่างการคำนวณค่า TF-IDF	10
4. การวิเคราะห์ข้อมูล	21
5. คำอธิบาย Use case คุณมีการใช้งาน	24
6. คำอธิบาย Use case ดูรายชื่อบริษัททั้งหมด	24
7. คำอธิบาย Use case ดูรายชื่อบริษัทในกลุ่มทั้งหมด	25
8. คำอธิบาย Use case ดูข้อมูลบริษัท	25
9. คำอธิบาย Use case คนหาบริษัทด้วยความสนใจ	25
10. คำอธิบาย Use case แก้ไขคุณมีการใช้งาน	26
11. คำอธิบาย Use case เพิ่ม ลบ แก้ไขข้อมูลบริษัท	26
12. อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้	27
13. อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบริษัทใหม่	28
14. พจนานุกรมข้อมูลบริษัท	31
15. แสดงการนับจำนวนบริษัทแต่ละประเภทในการจัดกลุ่ม	43
16. แสดงจำนวน ค่าเฉลี่ยความพึงพอใจต่อระบบ	57
17. ขอเสนอแนะของผู้ใช้	57

สารบัญภาพ

ภาพที่	หน้า
1. ตัวอย่างการทำ Word segmentation	8
2. ตัวอย่างการจัดกลุ่มข้อมูลด้วย K-Means โดยที่เครื่องหมายกากรบทลีเดงคือจุด Centroid ของแต่ละกลุ่มข้อมูล	11
3. กราฟที่แสดงจำนวนของผิดพลาดเพื่อหาจพนวนกลุ่มที่เหมาะสมที่สุด	11
4. คำนวณค่า Cosine similarity ด้วย scikit-learn	12
5. ตัวอย่างการส่งคำขอไปยัง API	13
6. การตอบกลับจาก API	13
7. ตัวอย่างโค้ดสำหรับการสร้าง Web API ด้วย fastAPI	14
8. ผลลัพธ์แสดงคำว่า Hello project จาก fastAPI	14
9. ตัวอย่างข้อมูลแบบ JSON	16
10. การเตรียมและวิเคราะห์ข้อมูล	22
11. การทำงานของระบบ	23
12. Use Case Diagram ระบบแนะนำบริษัทสำหรับฝึกงานตามความสนใจ	24
13. Sequence Diagram การคุยกันระหว่างความสนใจของผู้ใช้	27
14. Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบริษัทใหม่	28
15. Activity Diagram ของผู้ใช้งาน	29
16. Activity Diagram ของผู้ดูแลระบบ	30
17. ER Diagram ระบบแนะนำบริษัทสำหรับฝึกงานตามความสนใจ	31
18. หน้าแรก	32
19. หน้าเกี่ยวกับ	33
20. หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด	33
21. หน้าแสดงรายชื่อบริษัททั้งหมด	34
22. หน้าแสดงผลลัพธ์รายชื่อบริษัท	34
23. หน้าแสดงข้อมูลบริษัท	35
24. ตัวอย่างข้อมูลต้นฉบับ	37
25. ตัวอย่างการตัดคำโดยใช้ Engine newmm	37
26. ตัวอย่างการตัดคำโดยใช้ Engine longest	38
27. ตัวอย่างการตัดคำโดยใช้ Engine deepcut	38
28. ผลการวัดค่าความแม่นยำในการตัดคำของ Engine ในไลบรารี Pythainlp	39
29. ตัวอย่างการตัดคำและเรียงคำที่มีน้ำหนักมากที่สุด 5 อันดับ	39
30. ตาราง TF-IDF แสดงน้ำหนักของคำ	40
31. ตัวอย่างการตัดคำและลบ Stop word	40
32. การทำ Elbow method	41
33. จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม	41
34. จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม	42
35. จัดกลุ่มข้อมูลจำนวน 7 กลุ่ม	42

	๊
36. จัดกลุ่มข้อมูลจำนวน 6 กลุ่ม	43
37. ตัวอย่างการสุมข้อมูลของการจัดกลุ่มข้อมูลที่ 9 กลุ่ม	46
38. ตัวอย่างการสุมข้อมูลของการจัดกลุ่มข้อมูลที่ 8 กลุ่ม	46
39. ตัวอย่างการสุมข้อมูลของการจัดกลุ่มข้อมูลที่ 7 กลุ่ม	46
40. ตัวอย่างข้อมูลในกลุ่ม 0	48
41. ตัวอย่างข้อมูลในกลุ่ม 2	48
42. หน้าแรกเว็บไซต์ Intern-assistant	49
43. คนหาบริษัท	49
44. หน้าแสดงผลลัพธ์การค้นหา	50
45. หน้าเกี่ยวกับ	50
46. หน้าแสดงรายชื่อบริษัททั้งหมด	51
47. หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด	51
48. หน้ารายละเอียดบริษัท	52
49. ตัวอย่างการคำนวณค่า Cosine similarity	52
50. ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API	53
51. หน้า Repository ใน Github	63
52. ดาวน์โหลด Repository ด้วยคำสั่ง git clone	63
53. แสดงการบันทึกไฟล์ข้อมูลบริษัทในไฟล์เดียว	64
54. แสดงการเปิดไฟล์ clustering.py ในโปรแกรม Visual studio code	64
55. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k	65
56. แสดงการรันคำสั่งจัดกลุ่มข้อมูลใน Terminal	65
57. แสดงไฟล์ clustered_company.csv	65
58. แสดงหน้าเว็บไซต์ MongoDB	66
59. แสดงหน้าการจัดการ Cluster MongoDB	66
60. แสดงหน้าตั้งค่าและสร้าง Cluster MongoDB	67
61. แสดงหน้าสร้างปัญชีสำหรับจัดการฐานข้อมูล	67
62. แสดงหน้าเพิ่ม IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้	68
63. แสดงหน้าจัดการ Cluster MongoDB	68
64. แสดงการสร้างฐานข้อมูล MongoDB	69
65. แสดงหน้าต่างการสร้างฐานข้อมูลและ Collection	69
66. หน้าเว็บไซต์สำหรับโหลดโปรแกรม MongoDB compass	70
67. ตัวอย่างการเลือกตั้งค่าการดาวน์โหลดโปรแกรม MongoDB compass	70
68. หน้าต่างเลือกเชื่อมต่อ กับ Cluster	71
69. หน้าต่างข้อมูลการเชื่อมต่อ Cluster กับ MongoDB compass	71
70. หน้าต่างโปรแกรม MongoDB compass สำหรับเชื่อมต่อ Cluster	72
71. หน้าต่างโปรแกรมแสดงข้อมูลใน Collection	72
72. หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv	73
73. หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass	73
74. หน้าเว็บไซต์ Amazon Web Services	74
75. หน้าแสดงการเลือกสร้าง Instance ใหม่	74

	๊
76. หน้าแสดงการตั้งค่า Instance	75
77. ตัวอย่างการเชื่อมต่อเข้าไปยัง Instance	75
78. การดาวน์โหลดโปรเจคจาก Github ด้วยคำสั่ง git clone มา�ัง Instance และใช้คำสั่ง cd final_project เพื่อเข้าไปยังโฟลเดอร์	76
79. สร้างไฟล์ใหม่ชื่อ .env และสร้างตัวแปรเพื่อกีบค่าเชื่อมต่อฐานข้อมูล	76
80. ตัวอย่างการรัน Server Web API เพื่อคำนวณค่า Cosine similarity บน AWS	77
81. หน้าเว็บไซต์ Vercel	77
82. สร้างโปรเจคใหม่ใน Vercel	78
83. หน้าแสดงรายชื่อ Repository	78
84. หน้าการตั้งค่าโปรเจคก่อน Deploy	79
85. ตัวอย่างหน้าเว็บไซต์	79
86. แสดงการเปิดไฟล์ clustering.py ในโปรแกรม Visual studio code	81
87. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k	81
88. แสดงการรันคำสั่งจัดกลุ่มข้อมูลใน Terminal	82
89. แสดงไฟล์ clustered_company.csv	82
90. หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv	83
91. หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass	83
92. หน้าเว็บไซต์ intern-assistant.vercel.app	84
93. หน้าแสดงผลลัพธ์เมื่อคนหาบริษัท	84
94. หน้าแสดงข้อมูลบริษัท	85

บทที่ 1 บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในระบบการศึกษาระดับปริญญาตรีนั้นรายวิชาที่มีในการศึกษาปีสุดท้ายของหลักสูตรคือ รายวิชาที่จะต้องให้นักศึกษาแต่ละคนนั้นออกแบบ自己ทำงานที่สถานประกอบการต่างๆ ในช่วงระยะเวลาหนึ่ง ซึ่งเป็นสิ่งที่มีความสำคัญและความท้าทายเนื่องจากเป็นการที่นักศึกษาจะได้ทดลองทำงานจริง สถานการณ์จริง สถานที่จริง ในสถานประกอบการที่นักศึกษาได้เลือก

ดังนั้นการเลือกสถานประกอบการสำหรับฝึกงานจึงเป็นเรื่องที่ต้องให้ความสำคัญเป็นอย่างมากเนื่องจากหากสถานประกอบการที่เลือกนั้นรูปแบบธุรกิจหรืองานที่ทำนั้น ตรงกันกับความสามารถของนักศึกษา ก็จะเป็นผลดี เนื่องจากความรู้และทักษะที่ได้จากการทำงานนั้นสามารถนำไปต่อยอดและใช้งานจริงเมื่อจบการศึกษาและเข้าทำงาน แต่หากสถานประกอบการที่เลือกนั้น รูปแบบธุรกิจหรืองานที่ทำไม่ตรงกับความต้องการหรือทักษะของนักศึกษาอาจทำให้การฝึกงานนั้น ส้มเหลว หรืออาจไม่ได้ความรู้และทักษะที่ต้องการได้ และด้วยเทคโนโลยีปัญญาประดิษฐ์ในด้านของการประมวลผลภาษาธรรมชาตินี้พัฒนาการหัวข้อมาก ทั้งในแง่ของเทคนิค เครื่องมือ และองค์ความรู้ ทำให้เกิดตัวอย่างการนำข้อมูลมาประมวลผลที่มีประสิทธิภาพมากmay ในปัจจุบัน

ด้วยเหตุนี้จึงได้มีการเริ่มโครงการพัฒนาเว็บแอปพลิเคชันที่นักศึกษามีความสนใจในรูปแบบธุรกิจของสถานประกอบการนั้น เพื่อช่วยอำนวยความสะดวกแก่นักศึกษา มหาวิทยาลัย เทคโนโลยีราชมงคลล้านนา น่าสนใจ การค้นหาสถานประกอบการสำหรับออกฝึกงาน โดยการใช้วิธี ประมวลผลภาษาธรรมชาติเข้ามาช่วยจัดกลุ่มสถานประกอบการและเสนอรายชื่อสถานประกอบการ ที่เหมาะสมแก่นักศึกษาผ่านทางเว็บแอปพลิเคชัน

1.2 วัตถุประสงค์

1.2.1 เพื่อศึกษาและพัฒนาเว็บแอปพลิเคชันเพื่อแนะนำสถานประกอบการตามความสนใจของนักศึกษา

1.2.2 เพื่อศึกษาและวิเคราะห์ความสนใจของนักศึกษาในการหาสถานประกอบการสำหรับฝึกงาน

1.2.3 เพื่อศึกษาประสิทธิภาพการแบ่งกลุ่มของข้อมูลสถานประกอบการด้วยวิธีประมวลผลภาษาธรรมชาติ

1.3 ขอบเขตของโครงการ

1.3.1 ผู้ใช้สามารถค้นหาสถานประกอบการได้ด้วยรายละเอียดของงานหรือรูปแบบธุรกิจที่สนใจ

1.3.2 เว็บแอปพลิเคชันสามารถให้ข้อมูลบริษัทเพื่อการตัดสินใจในการฝึกงานได้

1.3.3 ใช้เทคโนโลยีการประมวลผลภาษาธรรมชาติ เพื่อจัดการคำและหาความคล้ายคลึงโดยใช้เทคโนโลยีการจัดกลุ่ม (K-Means) เพื่อจัดกลุ่มข้อมูล

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 ได้พัฒนาเว็บแอปพลิเคชันสำหรับการค้นหาสถานประกอบการสำหรับผู้คน
- 1.4.2 สามารถนำระบบประมวลผลภาษาชาติมาใช้ในการจุดกลุ่มข้อมูลได้อย่างแม่นยำ
- 1.4.3 เป็นช่องทางสำหรับการเลือกและหาข้อมูลของสถานประกอบการสำหรับออกฝึกงานของนักศึกษา

1.5 อุปกรณ์เครื่องมือที่ใช้ในโครงงาน

1.5.1 Programing language

1. Python
2. Javascript
3. HTML
4. CSS

1.5.2 Framework

1. Next.js
2. fastAPI

1.5.3 Database

1. MongoDB

1.5.4 Program

1. Microsoft Excel
2. Visual studio code
3. Postman
4. Firefox
5. Figma
6. Notion

1.5.5 Version control

1. Git
2. Github

1.5.6 Python library

1. Pythainlp
2. Matplotlibs
3. Pandas
4. Numpy
5. Sci-kit learn
6. nltk
7. python-dotenv

1.5.7 Javascript library

1. Tailwind CSS
2. cors
3. dotenv

- 4. sweetalert2
- 5. headlessui
- 6. heroicons
- 1.5.8 Global network
 - 1. Cloudflare
- 1.5.9 Cloud computing
 - 1. Amazon Web Services
 - 2. Vercel

1.6 นิยามศัพท์ที่ใช้ในโครงงาน

1.6.1 การประมวลผลภาษาธรรมชาติ (Natural language processing:NLP) เป็นเทคนิคแขนงหนึ่งในศาสตร์ของเทคโนโลยีปัญญาประดิษฐ์ ซึ่งเป็นการทำให้คอมพิวเตอร์เข้าใจ ตีความ และสื่อสารภาษาของมนุษย์ได้

1.6.2 การจัดกลุ่มข้อมูล (Clustering) หมายถึง เป็น Machine learning model ชนิดหนึ่งที่อยู่ในประเภท Unsupervised คือเป็นการที่นำข้อมูลเข้าไปให้ Model ประมวลผลโดยที่ไม่ได้จำกัดค่าตอบไว้แต่ให้คอมพิวเตอร์ประมวลผลและกำหนดเองว่าค่าตอบควรจะเป็นลักษณะใดบ้าง

1.6.3 การตัดคำ (Word segmentation) หมายถึง ด้วยที่การเขียนภาษาไทยนั้นไม่มีการแยกคำด้วยการเว้นวรรคหรืออักษรเดียว หรือ ภาษาอื่นๆ ดังนั้นจึงจำเป็นต้องทำการตัดคำจากประโยคของมาเป็นคำ ๆ เพื่อให้นำไปประมวลหรือใช้งานต่อได้ด้วยอัลกอริทึมต่าง ๆ

1.6.4 การหาความคล้ายคลึง (Cosine similarity) ระหว่างเวกเตอร์เอ (Vector A) และเวกเตอร์บี (Vector B) ว่าไปพิศทางเดียวกันหรือไม่โดยการใช้สูตรของกฎสามเหลี่ยมเพื่อหาผลลัพธ์แล้วนำมาเปรียบเทียบกัน

1.6.5 ช่องทางสำหรับการสื่อสารกัน (Application programming interface:API) ระหว่างเครื่องแม่ข่าย (Server) และ เครื่องลูกข่าย (Client) สร้างขึ้นมาเพื่อเป็นตัวกลางให้โปรแกรม หรือผู้ใช้อื่นๆ ใช้ติดต่อสื่อสาร เชือต่อแลกเปลี่ยนข้อมูลกัน

1.6.6 การเช่าใช้ระบบคอมพิวเตอร์ (Cloud computing) และทรัพยากรแบบครบวงจรจากผู้ให้บริการต่าง ๆ เช่น Amazon, Google, Microsoft, Huawei โดยสามารถกำหนดรูปแบบของฮาร์ดแวร์และซอฟต์แวร์ที่ต้องการได้ มีให้บริการทั้งเครื่องแม่ข่าย(Server) ฐานข้อมูล(Database) การทดสอบระบบ(Testing) หรือแอปพลิชันสำเร็จรูปในหลายระบบปฏิบัติการ(Platform)

1.6.7 สถานประกอบการ หมายถึง บริษัทที่ประกอบอาชีพทางด้านศาสตร์ของคอมพิวเตอร์ และเทคโนโลยีต่าง ๆ

1.6.8 ตรงกับความต้องการ หมายถึง การนำความต้องการของผู้ใช้มาเปรียบเทียบกับข้อมูลและคำนวณความคล้ายคลึง

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาคนครัวเพื่อจัดทำโครงการทางวิทยาการคอมพิวเตอร์ ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ ผู้ศึกษาได้ศึกษาคนครัวเอกสารและงานวิจัยที่เกี่ยวข้องตามหัวข้อลำดับต่อไปนี้

2.1 ทฤษฎีและความรู้ที่เกี่ยวข้อง

- 2.1.1 ทฤษฎี การประมวลผลภาษาธรรมชาติ หรือ Natural language processing
- 2.1.2 ทฤษฎี การตัดคำในภาษาไทยหรือ Word segmentation
- 2.1.3 ทฤษฎี การสกัดใจความของข้อความด้วยเทคนิค TF-IDF
- 2.1.4 ทฤษฎี การจัดกลุ่มข้อความด้วยอัลกอริทึม K-Means
- 2.1.5 ทฤษฎี การหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow method
- 2.1.6 ทฤษฎี การคำนวณค่าความคล้ายคลึงด้วยเทคนิค Cosine similarity
- 2.1.7 ทฤษฎี การจัดการระบบคลาวด์ (Amazon web service)
- 2.1.8 ทฤษฎี API
- 2.1.9 ทฤษฎี Cloudflare
- 2.1.10 ทฤษฎี Cors
- 2.1.11 ทฤษฎี Fastapi
- 2.1.12 ทฤษฎี Git
- 2.1.13 ทฤษฎี Node.js
- 2.1.14 ทฤษฎี Matplotlibs
- 2.1.15 ทฤษฎี Mongodb
- 2.1.16 ทฤษฎี Next.js
- 2.1.17 ทฤษฎี Numpy
- 2.1.18 ทฤษฎี Pandas
- 2.1.19 ทฤษฎี Pythainlp
- 2.1.20 ทฤษฎี Scikit-learn
- 2.1.21 ทฤษฎี Vercel

2.2 งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีและความรู้ที่เกี่ยวข้อง

- 2.1.1 ทฤษฎี การประมวลผลภาษาธรรมชาติ หรือ Natural language processing

การประมวลผลภาษาธรรมชาติ (Natural Language Processing:NLP) หรือภาษาของมนุษย์ที่ใช้สื่อสารกัน เป็นเทคนิคนึงในเทคโนโลยีปัญญาประดิษฐ์ ที่จะทำให้คอมพิวเตอร์สามารถเข้าใจและเรียนรู้ ประมวลผลภาษาของมนุษย์ได้ ในด้านของการวิเคราะห์ภาษาศาสตร์ การตีความจากบทความ หรือการทั้งการแปลภาษา NLP นั้นจำเป็นต้องใช้ความรู้จากหลาย ๆ ศาสตร์เข้ามา เช่น Mathematics, Linguistics, Psychology เพื่อพัฒนาประสิทธิภาพการทำงานและความฉลาดของคอมพิวเตอร์

จุดเริ่มต้นของ NLP นั้นมีมาตั้งแต่ประมาณปี 1950–1980 ในยุคหนึ่งวิธีการที่จะให้คอมพิวเตอร์เข้าใจภาษาของมนุษย์นั้นใช้ “Rule-based” เป็นการใช้ if-else ในโปรแกรมที่ตั้งไว้ตามคำที่กำหนด และในต่อมาประมาณปี 1981–2001 เริ่มมีการใช้ ML หรือ Machine learning ที่ใช้อัลกอริทึมในการประมวลผล เช่น Decision Tree เข้ามาช่วยในการประมวลผล และฝึกสอนคอมพิวเตอร์โดยข้อมูลที่เป็น Dataset ทำให้ความแม่นยำเพิ่มขึ้น และในยุคปัจจุบันยุคที่มี Deep Neural Network เนื่องจากปัจจุบันคอมพิวเตอร์มีความสามารถที่เพิ่มขึ้นและปริมาณข้อมูลนั้นมีมากขึ้นตาม ทำให้การใช้ Deep Neural Network มาสร้างโมเดลสำหรับการทำ NLP เป็นที่นิยมมากยกตัวอย่างเช่น word embeddings คือการหา semantic กับข้อความนั้นๆ

กระบวนการทำงานของ NLP นั้น มีประกอบไปด้วยหลายส่วนของการประมวลผลและใช้แปลความหมาย ประกอบด้วยดังนี้

1. Tokenization เป็นการตัดคำออกเป็นคำ ๆ เพื่อที่จะนำไปประมวลผลต่อตามรูปแบบของแต่ละภาษา
2. Parsing เป็นการระบุโครงสร้างของข้อความ
3. Lemmatization/stemming คือ การแปลงคำให้อยู่ในรูปแบบเดียวเดิม
4. Part-of-speech tagging คือ การอธิบายหรือการกำหนดว่าในแต่ละคำนั้นมีความหมาย หรือประเภทของคำเป็นอย่างไร
5. Language detection การตรวจสอบภาษาว่าเป็นภาษาอะไร
6. Identification of semantic relationships คือการระบุความสัมพันธ์ของคำต่าง ๆ ในประโยค

ปัจจุบัน NLP นั้นอยู่ในหลายรูปแบบรอบตัวถูกนำมาใช้ในหลาย ๆ ด้านทั้ง Digital marketing, ทางการแพทย์ การแปลงภาษา Chatbot และอื่น ๆ (ตายะ, 2022)

2.1.2 ทฤษฎี การตัดคำในภาษาไทย (Word segmentation)

การที่นำประโยชน์มาตัดออกเป็นคำ ๆ (Word segmentation) เนื่องจากในบางภาษา เช่น ภาษาไทยรูปแบบการเขียนนั้นไม่มีการเว้นวรรคของคำต่างจากภาษาอังกฤษที่ใช้การเว้นวรรคในแต่ละคำ ดังนั้นถ้าจะทำ NLP ที่เป็นภาษาไทยนั้นจำเป็นต้องทำ Word segmentation เพื่อให้ได้ชุดคำที่จะนำไปใช้งานต่อ ในปัจจุบันการทำ Word segmentation นั้นมีเครื่องมือให้เช่นอย่างภาษา Python library pythainlp, nltk หรือสามารถใช้บริการ web API ของ aiforthai

```
onze@Tinngrits-MacBook-Pro:~/Documents/final_project/report/demo
python main.py
['นอน', 'ตกลม', 'ดู', 'ดาว']
```

ภาพที่ 1 ตัวอย่างการทำ Word segmentation

จากการที่ 1 เป็นการทำ Word segmentation ด้วย Python library pythainlp จากคำว่า “นอนตากลมดูดาว” ได้ผลลัพธ์ออกมาเป็น นอน, ตากลม, ดู, ดาว (L, 2019)

2.1.3 ทฤษฎี การสัดส่วนความของข้อความด้วยเทคนิค TF-IDF

การทำ Term Frequency – Inverse Document Frequency (TF-IDF) เป็นเทคนิคที่พิจารณาองค์ประกอบของคำภาษาในประโยชน์ เทคนิคนี้มากจาก 2 องค์ประกอบต่อกันคือ Term Frequency (TF) และ Inverse Document Frequency (IDF) องค์ประกอบแรก Term Frequency (TF) นั้นหมายถึงการที่หาคำที่มีการใช้ซ้ำบ่อยที่สุดในเอกสารนั้น ๆ ซึ่งแสดงไปถึงว่าคำนั้นเป็นคำที่มีความสำคัญมากเอกสารนั้น วิธีคำนวนค่าความถี่ของคำใช้การนำจำนวนครั้งของคำที่ปรากฏในเอกสารมาหารด้วยจำนวนคำทั้งหมดในเอกสาร เช่น ต้องการหาค่าความถี่ของคำว่าเว็บไซต์ในเอกสาร (CHAKRIT, 2019)

$$TF(\text{ของคำคำนี้}) = \frac{\text{จำนวนของคำนั้นที่มีในเอกสาร}}{\text{จำนวนคำทั้งหมดที่มีในเอกสาร}}$$

ตาราง 1 ตัวอย่างการคำนวนค่า Term Frequency ที่จำนวนคำทั้งหมดเท่ากับ 7

คำ	จำนวนคำ	Term Frequency	ผลลัพธ์
เว็บไซต์	5	5 ÷ 7	0.71
หนังสือ	1	1 ÷ 7	0.14
ออนไลน์	2	2 ÷ 7	0.29
ขาย	2	2 ÷ 7	0.29
เข้าชม	1	1 ÷ 7	0.14
มือถือ	4	4 ÷ 7	0.57
และ	3	3 ÷ 7	0.43

จากตัวอย่างจะเห็นได้ว่าคำว่าเว็บไซต์ปรากฏบ่อยในเอกสารทำให้มีค่า Term Frequency สูง จึงเรียกได้ว่าเป็นคำสำคัญของเอกสาร แต่การใช้ค่า Term Frequency เพื่อหาใจความสำคัญเพียงอย่างเดียวนั้นยังไม่ดีพอ จึงต้องใช้องค์ประกอบ Inverse Document Frequency (IDF) เข้ามาช่วยเพิ่มเติม Inverse Document Frequency (IDF) หมายถึง การคำนวนหาจำนวนคำของคำโดยการนำคำสำคัญค้นหาจากหลาย ๆ เอกสารหากคำนั้นมีค่า Inverse Document Frequency (IDF) ต่ำแสดงว่าคำนั้นไม่ได้เป็นคำสำคัญของเอกสารทั้งหมด สมการที่ใช้คำนวนหาค่า Inverse Document Frequency (IDF)

$$IDF(\text{ของคำคำนี้}) = \log\left(\frac{\text{จำนวนเอกสารทั้งหมด}}{\text{จำนวนเอกสารที่มีคำนั้นปรากฏ}}\right)$$

ตาราง 2 ตัวอย่างการคำนวณค่า Inverse Document Frequency ที่จำนวนเอกสารเท่ากับ 10

คำ	จำนวนเอกสารที่ปรากฏ	Inverse Document Frequency	ผลลัพธ์
เก็บไซต์	5	$\log(10 \div 5)$	0.31
หนังสือ	2	$\log(10 \div 2)$	0.70
ออนไลน์	2	$\log(10 \div 2)$	0.70
ขาย	2	$\log(10 \div 2)$	0.70
เข้าชม	1	$\log(10 \div 1)$	1.00
มือถือ	3	$\log(10 \div 3)$	0.52
และ	2	$\log(10 \div 2)$	0.70

จากนั้นคำนวณค่า Term Frequency – Inverse Document Frequency (TF-IDF) ได้โดยสมการ

$$TF - IDF = TF \times IDF$$

ตาราง 3 ตัวอย่างการคำนวณค่า TF-IDF

คำ	TF	IDF	TF-IDF
เก็บไซต์	0.71	0.31	0.22
หนังสือ	0.14	0.70	0.10
ออนไลน์	0.29	0.70	0.20
ขาย	0.29	0.70	0.20
เข้าชม	0.14	1.00	0.14
มือถือ	0.57	0.52	0.30
และ	0.43	0.70	0.30

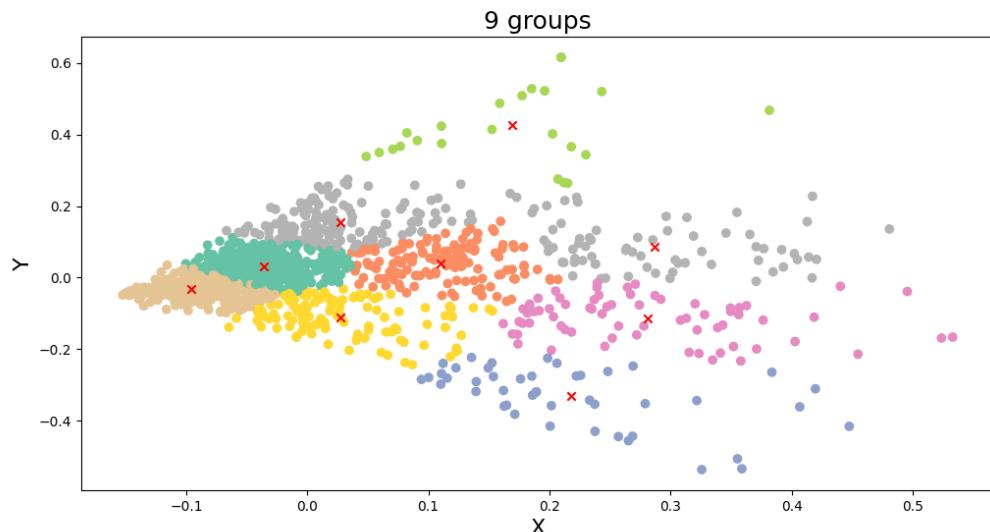
ดังตัวอย่างจะเห็นได้ว่าบางคำที่มีค่า TF-IDF สูงแต่ไม่ได้บ่งบอกถึงลักษณะของข้อความในเอกสาร เช่นคำว่า และ ซึ่งถือว่าเป็น Stop word ซึ่งเป็นคำที่ไม่สื่อความหมายโดยปกติแล้ว คำเหล่านี้มักถูกกรองออกก่อนที่จะมีการนำข้อมูลมาทำการประมวลผลทางภาษา自然ภาษาตัวอย่างจะเห็นว่าเมื่อคำนวณหาค่า Term Frequency – Inverse Document Frequency (TF-IDF) โดยที่ตัดคำที่ไม่มีความหมายหรือ Stop word ออกแล้วจะเหลือคำว่า มือถือ เก็บไซต์ ออนไลน์ เรียงลำดับความสำคัญจากมากไปน้อยตามลำดับ (Patipan, 2020)

2.1.4 ทฤษฎี การจัดกลุ่มข้อมูลความตัวอย่างอัลกอริทึม K-Means

K-Means เป็นวิธีการหนึ่งใน Data mining อยู่ในกลุ่มของ Unsupervised Learning คือการให้คอมพิวเตอร์เรียนรู้โดยไม่ต้องมีผู้สอน (Chakrit, 2018) เป็นอัลกอริทึมสำหรับการทำ Clustering Model เป็นการจัดกลุ่มข้อมูลด้วยการกำหนดจำนวนกลุ่มก่อนการทำ Clustering ซึ่งแทนด้วยค่า K จากนั้นคำนวณหาจุดกึ่งกลางของแต่ละกลุ่มเรียกว่าจุด Centroid ตามจำนวนกลุ่มที่กำหนดไว้ เว้นระยะห่างด้วยการคำนวณระยะห่างด้วยสมการ (Chakrit, ว่าด้วย k-means และการประยุกต์, 2018)

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

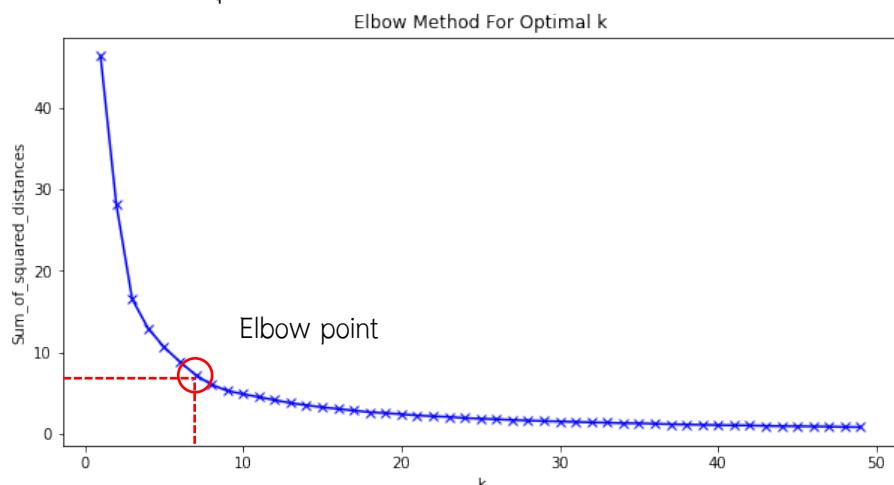
ทำการคำนวณและขยับจุด Centroid และหาค่าเฉลี่ยจนค่าเฉลี่ยไม่มีการเปลี่ยนแปลงจะได้จุดกึ่งกลางของข้อมูลในแต่ละกลุ่ม



ภาพที่ 2 ตัวอย่างการจัดกลุ่มข้อมูลด้วย K-Means โดยที่เครื่องหมายกาบบาทลีเดงคือจุด Centroid ของแต่ละกลุ่มข้อมูล

2.1.5 ทรัพยากรากฐานที่เหมาะสมสมด้วยวิธี Elbow method

Elbow method เป็นวิธีหนึ่งที่ใช้หาจำนวนของกลุ่มที่เหมาะสมสมด้วยการวัดขอผิดพลาด (Error measurement) ผลรวมระยะห่างระหว่างข้อมูลกับจุด Centroid เมื่อขอผิดพลาดน้อยลงความชันของเส้นโค้งจะแบนราบไปตามแกน X จนทำให้เกิดมุมลักษณะเหมือนกับขอศอกก็จะถือว่ามีจุดที่อยู่ตรงมุมขอศอกเป็นจำนวนของกลุ่มข้อมูลที่เหมาะสมสมดังในภาพตัวอย่างภาพที่ 3 จะเห็นได้ว่าจำนวนกลุ่มที่เหมาะสมสมคือ 4-5 กลุ่ม (Paul, 2021)



ภาพที่ 3 กราฟที่แสดงจำนวนขอผิดพลาดเพื่อหาจำนวนกลุ่มที่เหมาะสมที่สุด

2.1.6 ทฤษฎี การคำนวณค่าความคล้ายคลึงด้วยเทคนิค Cosine similarity

การวัดความเหมือนของ Vector 2 (Cosine Similarity) ว่าไปในทิศทางเดียวกันหรือไม่ โดยที่เป็นการตัดขนาด หรือ Magnitude ของ Vector ออกไปหาค่าได้จากการนี้

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

อธิบายโดยง่ายคือเป็นการวัดระยะห่างระหว่าง Object A และ Object B ว่ามีความคล้ายกันแค่ไหนยกตัวอย่างเปรียบเทียบระหว่างคำว่า “ยินดีที่ได้รู้จักรับ” และ “ยินดีที่ได้รู้จัค่ะ” ทำการตัดเพื่อหาคำทั้งหมดก่อนครับ คือ [“ยินดี”, “ที่”, “ได้”, “รู้จัก”, “ครับ”, “ค่ะ”] เราชะได้ Object A และ B เช่นเป็นชุดข้อมูล ได้ดังนี้ (Supalerk, 2020)

- A. “ยินดีที่ได้รู้จักรับ” = [1 , 1 , 1 , 1 , 1 , 0]
- B. “ยินดีที่ได้รู้จัค่ะ” = [1 , 1 , 1 , 1 , 0 , 1]

```
onze@Tinngrits-MacBook-Pro:~/documents/final_project/report/demo
python main.py
Cosine similarity: 0.8
~/documents/final_project/report/demo>
```

ภาพที่ 4 คำนวณค่า Cosine similarity ด้วย scikit-learn

2.1.7 ทฤษฎี การจัดการระบบคลาวด์ (Amazon web service)

AWS เป็นตัวย่อของ Amazon Web Services ซึ่งเป็นบริการบนระบบคลาวด์ ที่มีบริการหลากหลายมากกว่า 200 โซลูชัน ถูกใช้งานในธุรกิจและองค์กรทุกประเภทไม่ว่าจะเป็นบริษัท สตาร์ตอัป องค์กรขนาดใหญ่ ไปจนถึงหน่วยงานของรัฐ AWS ให้บริการโครงสร้างพื้นฐานด้านไอที การใช้บริการ Server และ Storage การสร้างและดูแลเว็บไซต์ ไปจนถึง ระบบอี-คอมเมิร์ซ การสร้างแอปพลิเคชัน การส่งเสริมการทำงานแบบ Remote Working การใช้ระบบ IoT เพื่อการสร้างนวัตกรรมใหม่ ๆ รวมถึงโซลูชันอื่น ๆ ในปัจจุบันนี้ AWS เป็นระบบประมวลผลบนคลาวด์ที่มีผู้ใช้บริการมากที่สุดในโลก เพราะได้รับความไว้วางใจจากผู้คนทั่วโลก เนื่องจาก AWS เป็นบริษัทในเครือของ Amazon เก็บไชต์ซื้อขายสินค้าออนไลน์ชื่อดังจากประเทศสหรัฐอเมริกา (CloudHM, 2022)

2.1.8 ทฤษฎี API

API ย่อมาจาก (Application Program Interface) ส่วนต่อประสานโปรแกรมประยุกต์ ในบริบทของ API คำว่า “Application” หมายถึงทุกซอฟต์แวร์ที่มีฟังก์ชันชัดเจน ส่วน “Interface” อาจถือเป็นสัญญาบริการระหว่างสองแอปพลิเคชัน ใช้สื่อสารกันโดยใช้คำขอ (Request) และการตอบกลับ (Response) ระหว่างเครื่องแม่ข่ายและแอปพลิเคชันอื่น ๆ API คือกลไกที่ช่วยให้ส่วนประกอบซอฟต์แวร์สองส่วนสามารถสื่อสารกันได้โดยใช้ชุดคำจำกัดความและproto-col ตัวอย่างเช่น ระบบซอฟต์แวร์ของสำนักพยากรณ์อากาศประกอบด้วยข้อมูลสภาพอากาศรายวัน (API คืออะไร)

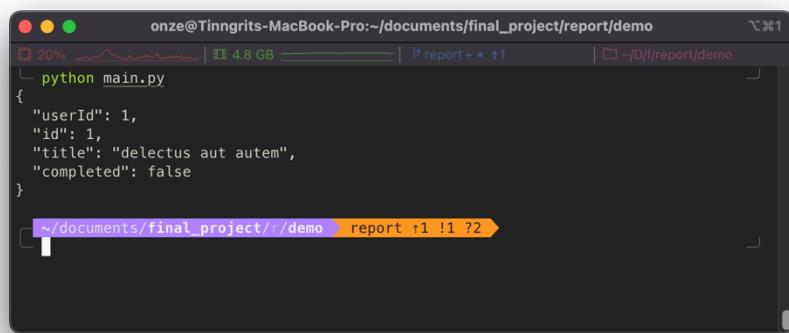


```

1 import requests
2
3 api_url = 'https://jsonplaceholder.typicode.com/todos/1'
4 response = requests.get(api_url)
5 print(response.text)

```

ภาพที่ 5 ตัวอย่างการส่งคำขอไปยัง API



```

onze@Tinngrits-MacBook-Pro:~/documents/final_project/report/demo
python main.py
{
  "userId": 1,
  "id": 1,
  "title": "delectus aut autem",
  "completed": false
}

```

ภาพที่ 6 การตอบกลับจาก API

2.1.9 ทฤษฎี Cloudflare

คลาวแฟร์ (Cloudflare) คือ Global Network ที่ถูกออกแบบมาเพื่อให้ทุกสิ่งของคุณที่เชื่อมต่อในเครือข่าย มีความปลอดภัย (Security) มีประสิทธิภาพ (Performance) และพร้อมใช้งาน (Availability) ซึ่ง Cloudflare จะทำหน้าที่เป็นตัวกลางระหว่างผู้เข้าใช้งานและ Server ที่เก็บข้อมูล โดยผู้เข้าใช้งานจะมาทั้งในรูปแบบของ Visitor, Crawlers & Bots และ Attackers แต่เมื่อใช้งาน Cloudflare การเข้าถึงทุกรูปแบบจะต้องผ่านระบบของ Cloudflare แทน โดย Cloudflare จะเข้ามาช่วยใน 3 เรื่องหลัก ๆ คือ (Cloudflare คืออะไร จะเข้ามาช่วยองค์กรของคุณได้อย่างไร?, 2021)

1. Web Application Firewall (WAF) ป้องกันการโจมตีเว็บไซต์ในรูปแบบ Cloud Security โดย WAF จะช่วยกัน HTTP/HTTPS Traffic ที่เป็นอันตรายออกโดยอัตโนมัติ เช่น Code Injection, Cross-Site-Scripting และ Sensitive Data Exposure

2. Distributed Denial-of-Service (DDoS) คือการโจมตีโดยการส่ง Traffic ปริมาณมากไปยังเว็บไซต์ เพื่อขัดขวางความสามารถในการให้บริการ หรือทำให้ไม่สามารถใช้งานได้ โดย Cloudflare จะเข้ามารับการโจมตีดังกล่าวแทนเว็บไซต์

3. Content Delivery Network (CDN) คือ การกระจายเนื้อหาออกไปตาม Server จุดต่าง ๆ หากมี Traffic ระบบจะส่งข้อมูลโดยใช้ Server ที่อยู่ใกล้ที่สุด โดย Cloudflare มี POPs ในไทยมากถึง 6 POPs และมากกว่า 200 POPs ทั่วโลก ช่วยให้เว็บไซต์สามารถใช้งานได้อย่างรวดเร็ว และเสถียร

2.1.10 ทฤษฎี Cors

การอนุญาตการแบ่งปันข้อมูลกัน (Cross-Origin Resource Sharing:CORS) เป็นกลไกที่ใช้เพิ่มเติมเพื่อให้บรรลุเป้าหมายนี้ในการเข้าถึงทรัพยากรที่เลือกจากเซิร์ฟเวอร์บนโดเมนอื่นมาแสดงบนหน้าเว็บบรรลุเป้าหมายนี้ได้ คอมพิวเตอร์แต่ละเครื่องต้องมี Protocol ที่เหมือนกัน ซึ่งจะสืบสานกันรุ่นเรื่อง เว็บบรรลุเป้าหมายนี้ได้ HTTP request เมื่อต้องการขอข้อมูลข้ามโดเมนหรือ port ที่ต่างกัน และต้องทำตามข้อตกลงการสื่อสาร (Protocol) เพราะปัจจุบันเรามักจะแยกผัง Front-end และ Back-end ออกจากกันเป็นคนละโดเมน ด้วยเหตุผลเรื่องความปลอดภัยของ Browsers HTTP การอนุญาตให้เข้าถึงแหล่งข้อมูลจะต้องอยู่ในโดเมนเดียวกันเท่านั้น เว้นแต่ว่าแหล่งข้อมูลนั้นจะอนุญาตให้โดเมนของ Browsers สามารถเข้าถึงข้อมูลเหล่านั้นได้ (TAeng Trirong, 2017)

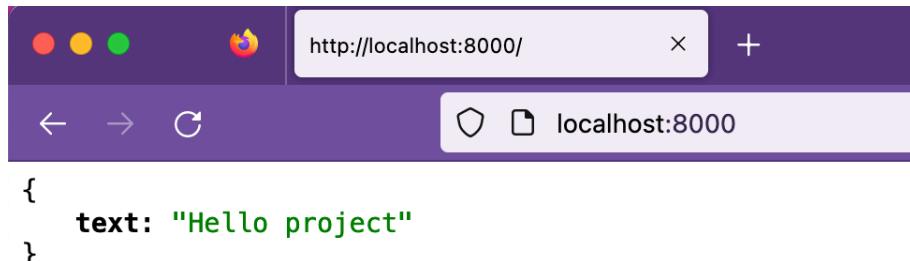
2.1.11 ทฤษฎี Fastapi

เฟรมเวิร์คสำหรับพัฒนาส่วนต่อประสานเครื่องแม่ข่ายกับเครื่องลูกข่ายด้วยภาษา Python (fastAPI) ถูกออกแบบมาให้ง่ายต่อการพัฒนา และสามารถที่จะสร้าง API ขึ้นมาได้อย่างรวดเร็ว โดยประสิทธิภาพการทำงานนั้นเร็ว FastAPI นั้นรองรับการทำงานแบบ Asynchronous และเมื่อเจ้าตัว Uvicorn เป็นตัว run server ข้อดีของการใช้งาน fastAPI คือ (Natakorn, 2021)

1. มีความเร็วของการทำงานเท่า Node.js และ Go
2. รูปแบบการเขียนฟังก์ชันต่าง ๆ เข้าใจง่ายต่อการศึกษา
3. ง่ายต่อการใช้งานและพัฒนาต่อ

```
● ● ●
1 from typing import Union
2 from fastapi import FastAPI
3
4 app = FastAPI()
5
6 @app.get('/')
7 def read_root():
8     return {'text': "Hello project"}
```

ภาพที่ 7 ตัวอย่างโค้ดสำหรับการสร้าง Web API ด้วย fastAPI



ภาพที่ 8 ผลลัพธ์แสดงคำว่า Hello project จาก fastAPI

2.1.12 ทฤษฎี Git

Git คือ Version Control ที่ถูกพัฒนาขึ้นมาเพื่อใช้ในกระบวนการพัฒนาซอฟต์แวร์ แปลให้เข้าใจแบบง่าย ๆ คือ ระบบที่ถูกพัฒนาขึ้นมาเพื่อใช้สำหรับการติดตาม ตรวจสอบ การพัฒนา แก้ไข ซอร์ซโคเด็ค ซอร์ซไฟล์ ต่าง ๆ ในขั้นตอนการพัฒนา ที่สามารถตรวจสอบได้ทุกด้านกันจร ทุกบรรทัด ทุกไฟล์ ที่มีการแก้ไข ใครเป็นคนแก้ไข และแก้ไข ณ วันที่เท่าไหร่

ระบบการทำงานของ Git ไม่ได้อยู่แค่การตรวจสอบการแก้ไขเท่านั้น ยังสามารถรวมการแก้ไขทั้งหมดเข้าด้วยกันได้อย่างชัมฉลาด เราเรียกขั้นตอนนี้ว่า CI (Continuous Integration) และ ในปัจจุบัน Git VCS (Version Control System) มีการรวมมิตร์ที่ทำให้นักพัฒนาทำงานได้สะดวกมากขึ้น สามารถทำงานได้ตั้งแต่ขั้นตอนการพัฒนา ไปจนถึงการ Deploy งานเข้าสู่งานบน Server เราเรียกขั้นตอนนี้ว่า CD (Continuous Deployment) รูปแบบการใช้งานของ Git มีด้วยกัน 2 รูปแบบ คือ (codebee, 2020)

- ใช้งานผ่าน Git Command Line (ใช้งานผ่านการพิมพ์คำสั่งด้วยตัวหนังสือ)
- ใช้งานผ่านโปรแกรม Git GUI (ใช้งานผ่านโปรแกรมล้ำเร็วๆ)

2.1.13 ทฤษฎี Node.js

Node.js คือสภาพแวดล้อมการทำงานของภาษา JavaScript นอกเว็บเบราว์เซอร์ที่ทำงานด้วย V8 engine นั้นหมายความว่าเราสามารถใช้ Node.js ในการพัฒนาแอพพลิเคชันแบบ Command line และพลิเคชัน Desktop หรือแม้แต่เว็บเบราว์เซอร์ได้ โดยที่ Node.js จะมี APIs ที่เราสามารถใช้สำหรับทำงานกับระบบปฏิบัติการ เช่น การรับค่าและการแสดงผล การอ่านเขียนไฟล์ และการทำงานกับเน็ตเวิร์ก เป็นต้น

Node.js ถูกพัฒนาและทำงานด้วย Chrome V8 engine สำหรับคอมโพล์ภาษา JavaScript ให้เป็นภาษาเครื่องด้วยการคอมโพล์แบบ Just-in-time (JIT) เพื่อเพิ่มประสิทธิภาพการทำงานของภาษา JavaScript จากที่แต่เดิมมันเป็นภาษาที่มีการทำงานแบบ Interpreted Node.js เป็นโปรแกรมที่สามารถใช้ได้ทั้งบน Windows, Linux และ Mac OS X นั้นหมายความว่าคุณสามารถเขียนโปรแกรมในภาษา JavaScript และนำไปรันได้ทุกระบบปฏิบัติการที่สนับสนุนโดย Node.js นี้เป็นแนวคิดของการเขียนครั้งเดียวแต่ทำงานได้ทุกที่ (Write once, run anywhere) ขอตีอีกอย่างหนึ่งใน การใช้ภาษา JavaScript ของ Node.js คือทำให้การพัฒนาเว็บไซต์ทำได้อย่างขึ้นสำหรับนักพัฒนา เนื่องจากเราสามารถใช้ภาษา JavaScript สำหรับทั้ง Front-end และ Back-end ได้โดยไม่ต้องศึกษาภาษาเฉพาะในแต่ละด้าน ตัวอย่างของการพัฒนาเว็บไซต์ในรูปแบบนี้ เช่น React.js ซึ่งเป็น ไลบรารีโดย Facebook (ทำความรู้จักกับ Node.js, 2021)

2.1.14 ทฤษฎี Matplotlibs

Matplotlib เป็นโมดูลที่เป็นพื้นฐานของ Python สำหรับการวาดกราฟจากข้อมูลซึ่งจำเป็นมากสำหรับงานทางด้าน Data Analysis, Science, Engineering เป็นตัวช่วยในการวิเคราะห์ข้อมูลโดยใช้รูปแบบของกราฟตัวอย่างประเทกราฟที่มีให้ใช้ 1.Scatter 2.Bar 3.Stem 4.Step และอื่น ๆ (หัด Python สำหรับคนเป็น Excel : ตอนที่ 8 – การสร้างกราฟด้วย Matplotlib)

2.1.15 ทฤษฎี Mongodb

MongoDB เป็น open-source document database โดยเป็นฐานข้อมูลแบบ NoSQL คือ ไม่มี relation (ความสัมพันธ์) ของตารางแบบ SQL ทั่วไป แต่จะเก็บข้อมูลเป็นแบบ JSON (JavaScript Object Notation) แทน การบันทึกข้อมูลทุกๆ record ใน MongoDB เราจะเรียกมันว่า Document ซึ่งจะเก็บค่าเป็น key และ value จะเห็นว่ามันก็คือ JSON นั่นแหลก (Chai, 2015) ตัวอย่างเช่น

```

2 [
3   {
4     "id": "63cced5afad0825781cde6697",
5     "title": "et autem facere repellat provident occaecati excepturi optio reprehenderit",
6     "body": "quia et suscipit nisi ut quas totam nostrum rerum est autem sunt rem eveniet architecto"
7   }
8 ]
9 [
10   {
11     "id": "63cced5afad0825781cde6697",
12     "title": "qui est esse",
13     "body": "et rerum temporis vitae et qui sint nihil reprehenderit dolor beatae ea dolores neque infugiat blanditiis voluptate porro vel nihil molestiae ut reiciendis qui aperiam non debitis possimus qui neque nisi nulla"
14   }
15 ]
16 ]
17 ]
18 ]
19 ]
20 ]
21 ]
22 ]
23 ]
24 ]
25 ]
26 ]
27 ]
28 ]
29 ]
30 ]
31 ]
32 ]

```

ภาพที่ 9 ตัวอย่างข้อมูลแบบ JSON

ข้อดีของตัวนี้เราจะพูดในหัวข้อถัดไป โดยหลักๆ ก็คือเหมาะสมกับองค์กรที่อยากจะเข้าถึงข้อมูลโดยอย่างรวดเร็ว อีกทั้งยังเหมาะสมกับการทำ Big Data และอื่น ๆ ดังนี้ (PLC, 2022)

1. ตัว MongoDB สามารถที่จะสร้างเป็น Cluster เพื่อที่จะตอบสนองของคำว่า High Availability (HA) ได้ ซึ่งเราอาจจะเลือก Region ที่เราอยากจะ Deploy บน Cloud Provider นั้นๆ ได้
2. ความรวดเร็วในการเข้าถึงข้อมูล เพราะว่าตัว Database ของมันไม่มี Schema ซึ่งจะต่างกับ SQL โดยพวknั้นจะอิงจากฐานข้อมูลที่มาจากการ Table
3. สามารถทำ Auto Scale ได้ไม่ว่าจะมีการใช้งานมากน้อยแค่ไหน ตามที่มันสามารถ Adapt กับ Environment นั้นๆ ได้
4. รองรับ Multiple Cloud Provider ซึ่งข้อดีข้อนี้มันจะทำให้ Database ของเรามี High Availability มากขึ้นโดยเราไม่จำเป็นที่จะต้องยึดติดกับ Cloud Provider เจ้าใดเจ้าหนึ่ง

2.1.16 ทฤษฎี Next.js

Next.js เป็น React Web Framework คล้าย ๆ กับ Create React App ที่ช่วยให้เราเขียนเว็บได้สะดวกขึ้น เพราะ Setup และ Config ให้เรียบง่ายครับถ้วน ยกตัวอย่างข้อดีของ Next.js เช่น (Pallop, 2017)

1. SSR (server-side rendering)
2. Hot rendering
3. Static HTML file exportable
4. Project Structure
5. Routing
6. Easy setting up & installation

สามารถทำเว็บไซต์ได้ทั้งแบบ static และ dynamic ซึ่งข้อดีของการเป็น server side rendering คือช่วยในเรื่อง SEO หรือ search engine optimization เพราะถ้าทำการ inspect เว็บไซต์ที่สร้างโดย Next.js จะเห็นว่า source จะเป็น html ส่วนใหญ่ ซึ่งทำให้ SEO คนผ่าน source เพื่อให้ได้ข้อมูลและจัดหมวดหมู่ได้ง่ายกว่า React ที่เป็น JavaScript มากกว่า ทำให้ Next.js เป็นที่นิยมในหลาย ๆ บริษัท นอกจากนี้ ข้อดีก็คือ render ได้เร็วกว่า React เพราะ Next.js มีสิ่งที่เรียกว่า get static path ซึ่งการสร้าง path แบบ static แบบเว็บไซต์ html โดยไม่ต้องทำการเชื่อมต่อกับ back end เพื่อให้ได้ data ยิ่งไปกว่านั้น Next.js สามารถรวมเข้ากับ backend ได้ง่ายๆ เพราะ Next.js มีสิ่งที่เรียกว่า API routes ใน การรับส่ง request ให้ใน folder ของ page จะมีอีก folder ที่เรียกว่า API ที่ถูกปฏิบัติเป็น endpoint แทนที่จะเป็น page ซึ่ง folder API นี้จะเป็นในส่วนหนึ่งของ server-side เท่านั้น ทำให้ไม่เพิ่ม size ของ client side (frevation, 2021)

2.1.17 ทฤษฎี Numpy

Numpy เป็น Library(ไลบรารี่) ที่รู้จัก และเป็นที่นิยมใช้ในการคำนวณ เช่น ใช้คำนวณ Matrix หรือ คำนวณกับ Array ในงาน Data Science, Data analytics และในการทำ Machine Learning (การเรียนรู้ของเครื่องจักร) หรือ Deep Learning (ดีพ เลิฟนิ่ง) ก็ยังคงต้องใช้ Numpy ออยู่ดี Numpy เป็น Library พื้นฐานที่ใช้คำนวณทางคณิตศาสตร์ด้วยภาษา Python สามารถคำนวณ หรือ ดำเนินการทำตระกูลใน Array หลายมิติ หรือ Matrix ได้อย่างรวดเร็ว เพราะ Library เขียนด้วยภาษา C ที่ Compile ไว้แล้ว (mindphp, Numpy คืออะไร)

Numpy นั้นได้แรงบันดาลใจมาจาก MATLAB ดังนั้นผู้ที่มีประสบการณ์ด้าน MATLAB อยู่แล้วจะทำความเข้าใจ Numpy ได้ไม่ยาก โดยหลักการของ คือการนิยามตัวแปร array หลายมิติ ที่เราคุ้นเคยในคณิตศาสตร์ อาทิ เช่น เวกเตอร์ (1 มิติ) เมटริกซ์ (2 มิติ) เทนเซอร์ (3 มิติขึ้นไป) เป็นต้น และ operations ของมัน ในการทำความเข้าใจ Numpy นั้นเพื่อนๆ ควรมีความรู้พื้นฐาน Linear algebra พวก vector / matrix ในระดับหนึ่ง (JUNG, 2019)

2.1.18 ทฤษฎี Pandas

pandas คือ หนึ่งใน Library สำคัญของภาษา Python เริ่มพัฒนาโดย Wes McKinney นักพัฒนาซอฟต์แวร์ชาวอเมริกัน ปัจจุบัน pandas เป็น open source ให้ทุกคนสามารถใช้ได้แบบฟรี pandas มาจากคำว่า Panel Data (ชุดข้อมูลหลายมิติ) มีจุดเด่นด้านการวิเคราะห์ข้อมูล (Data Analysis) และการทำความสะอาด (Data Cleaning) ซึ่งเป็น Process ที่สำคัญมากในการทำงานกับข้อมูล pandas มีความสามารถในการจัดการ และวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพตั้งแต่ข้อมูลขนาดเล็กไปจนถึงข้อมูลขนาดใหญ่ ทำให้ pandas ตอบโจทย์งานในยุคที่ข้อมูลมีขนาดใหญ่มากขึ้น เรื่อยๆ ได้ไม่ปัญหาติดขัดเหมือนกับ Spreadsheets อื่นๆ (เช่น Excel หรือ Google Sheets ซึ่งจะทำงานได้ช้าลงหากข้อมูลมีขนาดใหญ่ขึ้น) ขั้นตอนการเตรียมข้อมูลนั้นมีความสำคัญมาก และ Data Scientist อาจจะใช้เวลาส่วนใหญ่หมดไปกับขั้นตอนนี้ เพราะหากข้อมูลที่เตรียมไม่ได้มีประสิทธิภาพ การนำ Insights ไปใช้งาน หรือนำข้อมูลไปสร้างโมเดล ย่อมทำให้ได้ข้อมูลที่ไม่น่าเชื่อถือนั่นเอง

นอกจากนี้ เมื่อเปรียบเทียบกับ Tools วิเคราะห์ข้อมูลอื่นๆ อย่าง Excel หรือ Google Sheets อาจไม่ตอบโจทย์เต็มที่หากต้องการเชื่อมต่อกับแหล่งข้อมูลบางประเภท หรือทำ Automation (ระบบจัดการอัตโนมัติ) ในขณะที่ pandas ซึ่งเป็นส่วนหนึ่งของ Python นั้นสามารถใช้การเขียนโค้ดเพื่อปรับแต่ง หรือเชื่อมต่อกับโปรแกรมอื่นๆ ได้สะดวก (Panchart, 2021)

2.1.19 ทฤษฎี Pythainlp

ในส่วนของ Module pythainlp ก็เป็นเหมือนกันตัว library ที่รวมคำสั่งเกี่ยวกับที่เกี่ยวกับภาษาไทยใน Python ซึ่งก็เป็นตัวช่วยให้การทำงานเกี่ยวกับตัวของภาษาไทย มีประสิทธิภาพและสะดวกมากขึ้น ในการทำงานของ pythainlp ก็จะมีการทำงาน เช่น การตัดคำ การแปลไทยเป็นอังกฤษ และการเข้าถึงรหัส Soundex และยังมีการทำงานที่เกี่ยวกับตัวของภาษาไทย มีการแสดงผล เกี่ยวกับเช็ตของภาษาไทยทั้งหมด ยังมีในส่วนของการแยกคำ เช่นเป็นส่วนของพยัญชนะ สาร วรรณยุกต์ เป็นต้น ยังมีในส่วนของเลขไทย มีการเช็คว่าเป็นคำภาษาไทยรึไม่ มีการวนับตัวอักษร ว่าเป็นภาษาไทยกี่เบอร์เซ็นต์ และยังมีส่วนของการแสดงคำอ่านที่เป็นพวงเวลาและยังมีการจัดเรียงคำใน

List ให้เรียงกันเป็นลำดับได้ เป็นต้น (mindphp, 2022) นอกจานี้ยังใช้สำหรับประมวลผลข้อมูลความ และการวิเคราะห์ทางภาษา คล้ายกับ NLTK แต่ใช้กับภาษาไทยโดยเฉพาะ มีฟังก์ชันการทำงานที่หลากหลาย เช่น Character Set อักษรไทย คำไทย, เรียงคำภาษาไทย, Stop Words ภาษาไทย, ตัดคำภาษาไทย, วิเคราะห์ชนิดของคำทางไวยากรณ์, ตรวจตัวสะกด แก้คำผิด และอีกมากmany (Surapong, 2020)

2.1.20 ทฤษฎี Scikit-learn

Scikit-learn (เรียกอีกอย่างว่า sklearn) นำเสนอแบบจำลองทางสถิติและการเรียนรู้ของเครื่องที่หลากหลาย แตกต่างจากโมดูลส่วนใหญ่ sklearn ได้รับการพัฒนาใน Python มากกว่า C แม้จะได้รับการพัฒนาใน Python ก็ตาม ประสิทธิภาพของ sklearn นั้นถูกกำหนดให้ใช้ NumPy สำหรับการทำนายการพิเศษเชิงเด่นและอัลกอริتمที่มีประสิทธิภาพสูง (เจร์, 2021)

Scikit-Learn ถูกสร้างขึ้นโดยเป็นส่วนหนึ่งของโครงการ Summer of Code ของ Google และทำให้ชีวิตของนักวิทยาศาสตร์ข้อมูลที่มี Python เป็นศูนย์กลางนับล้านทั่วโลกง่ายขึ้น ส่วนนี้ของชีวิตมุ่งเน้นไปที่การนำเสนอไลบรารีและมุ่งเน้นไปที่องค์ประกอบเดียว นั่นคือการแปลงชุดข้อมูล ซึ่งเป็นขั้นตอนสำคัญและสำคัญที่ต้องทำก่อนพัฒนาแบบจำลองการทำนาย Scikit-learn เป็นแพ็คเกจ Python โอเพ่นซอร์สพร้อมการวิเคราะห์ข้อมูลที่ซับซ้อนและคุณสมบัติการชุด มันมาพร้อมกับอัลกอริทึมในตัวมากมายที่จะช่วยให้คุณได้รับประโยชน์สูงสุดจากโครงการวิทยาศาสตร์ข้อมูลของคุณ ห้องสมุด Scikit-learn ใช้วิธีดังต่อไปนี้

1. Classification
2. Regression
3. Clustering
4. Dimensionality reduction
5. Model selection
6. Preprocessing

2.1.21 ทฤษฎี Vercel

Vercel คือ Cloud Platform ที่ให้บริการทำ Static Hosting Website ต่างๆ และสามารถทำ Serverless Functions บน Cloud รวมทั้งยังสามารถ Integrate และสร้าง Workflow ผ่าน GitHub เพื่อทำ Automated Deployment โค้ดของคุณได้อย่างง่าย Vercel Inc. เดิมชื่อ Zeit เป็นแพลตฟอร์มคลาวด์ของอเมริกาในฐานะบริษัทผู้ให้บริการ บริษัทรักษาการของการพัฒนาเว็บไซต์ Next.js สถาปัตยกรรมของ Vercel สร้างขึ้นจาก Jamstack และการจัดการการปรับใช้ผ่านที่เก็บ Git Vercel เป็นสมาชิกของ MACH Alliance (Huangsri, 2021)

2.2 งานวิจัยที่เกี่ยวข้อง

จักรินทร์ สันติรัตนภักดี และศุภกฤษฐ์ นิวัฒนาภูล ศึกษาเรื่อง การออกแบบและพัฒนากระบวนการจำแนกชั้นของเรียนรถโดยสารสาธารณะเพื่อติดแท็กปัญหาการให้บริการ ของคุณภาพส่วนกลางชั้นกรุงเทพ (ขสมก.) มีช่องทางในการร้องเรียนรถโดยสารสาธารณะผ่านเว็บบอร์ด ที่ผู้ใช้งานสามารถแสดงความคิดเห็นได้อย่างอิสระ ผู้วิจัยจึงออกแบบและพัฒนากระบวนการจำแนกชั้นของเรียนรถโดยสารสาธารณะ จากชั้นของเรียนผ่านเว็บบอร์ดขององค์กรขนส่งมวลชนกรุงเทพด้วยกระบวนการตัดคำภาษาไทยโดยใช้พจนานุกรม แล้วคัดเลือกคำศัพท์ด้วยการวิเคราะห์น้ำหนักของคำมาสร้างเป็นคลังคำศัพท์ แบ่งเป็น 4 คลาส ได้แก่ คลาสการขับขี่ คลาสผู้ขับขี่ และพนักงานผู้ให้บริการ คลาสภายนอกและอุปกรณ์ให้บริการ และคลาสวีลาและการเดินรถโดยใช้มือถือการตัดคำภาษาไทย (Thai Word Segmentation) ด้วย ขนาดความทั่วไปซึ่งอยู่ในรูปแบบประโยคมาแบ่งออกเป็นค่าหรือคุณลักษณะ (Term/Feature) เพื่อแยกส่วนของข้อความออกจากกันก่อนนำไปประมวลผลในขั้นต่อไป แบ่งตามกระบวนการทำงานออกเป็น 3 กลุ่ม ได้แก่ 1) การตัดคำโดยใช้กฎ (Rule-Based Approach) 2) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) 3) การตัดคำโดยใช้คลังคำศัพท์ (Corpus-Based Approach) จากการทดลองพบว่าอัลกอริทึมโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวมสูงที่สุด (จักรินทร์ สันติรัตนภักดี, 2021)

วุฒิชัย วิเชียรไชย ศึกษาเรื่อง การเปรียบเทียบวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างการเขียนกับโครงสร้างพยานค์ งานวิจัยนี้นำเสนอการแบ่งแยกคำภาษาไทยโดยเทียบกับโครงสร้างการเขียนของภาษาไทยและอัลกอริทึมการแบ่งแยกคำภาษาไทยโดยโครงสร้างพยานค์ เพื่อศึกษาและเปรียบเทียบวิธีการประมวลผลของการแบ่งแยกคำภาษาไทยและประสิทธิภาพความถูกต้องของอัลกอริทึม โดยสามารถแบ่งงานวิจัยในการแบ่งแยกคำภาษาไทยได้เป็นดังนี้คือวิธีการใช้กฎ (Rule base approach) วิธีการใช้อัลกอริทึม (Algorithm ap-proach) วิธีการใช้พจนานุกรม (Dictionary base approach) และวิธีการใช้คลังข้อความ (Corpus based approach) ผู้วิจัยจึงได้เสนอวิธีการแบ่งแยกคำภาษาไทย โดยใช้โครงสร้างการเขียนภาษาไทยเพื่อแก้ไขลดพื้นที่ในการจัดเก็บคำศัพท์ในพจนานุกรม และวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างพยานค์เพื่อลดการสิ้นเปลืองพื้นที่ในการจัดเก็บพจนานุกรม ยกตัวอย่างการแบ่งแยกคำและพยานค์ของคำว่า “ประเทศไทย” จะสามารถแบ่งแยกคำได้เป็น “ประเทศไทย” และแบ่งพยานค์ได้เป็น “ประเทศไทย” จากผลลัพธ์ในการแบ่งแยกคำนั้นยังขาดความถูกต้องในการแบ่งแยกคำซึ่งสามารถพัฒนาแนวคิดในการศึกษาและสร้างกฎเพื่อแบ่งแยกคำให้ถูกต้องมากยิ่งขึ้น (วุฒิชัย, 2013)

ปราณี พึงวิชา อนาคต ทับเที่ยง และธัญญา สัตยาภิธาน (2019) ศึกษาการแบ่งกลุ่มพุทธิกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ ผู้วิจัยเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจำนวน 400 คน ทำการวิเคราะห์แบ่งกลุ่ม ผู้บริโภคด้วยวิธี K-Means Clustering เป็น 2 กลุ่มซึ่งมีลักษณะเฉพาะในแต่ละกลุ่ม จากการวิเคราะห์ความแตกต่างระหว่างกลุ่มของทัศนคติต่อ้านพุทธิกรรมการซื้อและด้านส่วนประสมทางการตลาดที่มีผลต่อการตัดสินใจซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์เมื่อออยู่ต่างกลุ่มกัน โดยการวิเคราะห์ความแปรปรวนทางเดียว(One-way ANOVA) พบว่า ด้านพุทธิกรรมการซื้อทั้ง 2 กลุ่ม มีความถี่ในการซื้อต่างๆ ไม่แตกต่างกัน จากการวิเคราะห์แบ่งกลุ่มผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ด้วยวิธี K-mean clustering สามารถจำแนกเป็น 2 กลุ่มโดยแต่ละกลุ่มมีลักษณะเฉพาะดังนี้ กลุ่มที่ 1 : กลุ่มกระเปาหนักจ่ายได้ถ้าชอบ ไม่ค่อยชอบออกสื่อ ลักษณะด้านประชากรศาสตร์ โดยส่วนใหญ่เป็น

คนGeneration X เพศหญิงมากกว่าเพศชาย มีระดับการศึกษาสูงกว่าปริญญาตรี กลุ่มที่ 2 : กลุ่มวัยสะอ้อน ชอบออกสื่อ เชื่อน้อยแต่บ่อยครั้ง ลักษณะด้านประชากรศาสตร์ โดยส่วนใหญ่เป็นคน Generation Y เพศหญิงมากกว่าเพศชาย ส่วนใหญ่มีระดับการศึกษาสูงแต่จะน้อยกว่ากลุ่ม 1 โดยมีระดับปริญญาตรีมากที่สุดโดยส่วนใหญ่เป็นพนักงานบริษัทเอกชน (ปาราณี พึงวิชา, 2019)

ธงชัย คล้ายคลึง วุฒิชัย สง่างาม กิตติวงศ์ สุธรรมโน และพันธุ์พงศ์อภิชาตกุลศ (2019) ศึกษาเรื่อง เทคนิคการคัดเลือกกลุ่มหลอดรายอาหารสำหรับองรับแผนการติดตั้งระบบผลิตไฟฟ้า พลังงานแสงอาทิตย์บนหลังคาเพื่อเพิ่มค่าครรชนีประสิทธิภาพการใช้พลังงานไฟฟ้า บทความนี้ต้องการนำเสนอเทคนิคหรือวิธีการคัดเลือกกลุ่มหลอดในแต่ละอาคารที่มีความเหมาะสมสำหรับติดตั้งระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์กรณีที่ติดตั้งบนหลังคาของอาคารในศูนย์กลางมหาวิทยาลัยเทคโนโลยีราชมงคลชีสานครราชสีมา ในวิธีการของ K-Means Clustering เริ่มต้นด้วยการจัดแบ่งข้อมูลออกเป็น K กลุ่ม กำหนดจุดศูนย์กลางที่ใกล้มากที่สุด จากผลการวิเคราะห์การใช้พลังงานไฟฟ้าทั้งหมด 34 อาคารด้วยวิธีการ K-Mean Clustering ทำให้สามารถแยกแยะจัดกลุ่มหลอดรายอาหารได้อย่างมีประสิทธิภาพซึ่งแบ่งได้ 3 กลุ่มโดยกลุ่มที่ 3 จำนวน 19 อาคารนั้นเป็นกลุ่มอาคารที่มีความเหมาะสมทั้งด้านพฤติกรรมการใช้พลังงานไฟฟ้าและมีพื้นที่รองรับการติดตั้งระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์บนหลังคาได้ (Thongchai Klayklueng, 2019)

วรากา ประดิษฐ์กุล ปราลี มณีรัตน์ และ นิเวศ จิระวิชัยชัย (2021) ศึกษาเรื่อง ระบบแนะนำรายนต์ให้กับลูกค้าโดยการวิเคราะห์จากการอ้างอิงถึงพฤติกรรมของผู้ใช้ (Collaborative Filtering) กรณีศึกษาบริษัท โตโยต้า บัสส์ จำกัด ผู้วิจัยได้พัฒนาระบบแนะนำรายนต์ให้กับลูกค้าโดยการวิเคราะห์จากการอ้างอิงถึงพฤติกรรมของผู้ใช้ เพื่อช่วยให้ลูกค้าได้รับการแนะนำรุ่นรถยนต์ที่เหมาะสม ตรงตามความต้องการของลูกค้า ผู้วิจัยใช้อัลกอริทึมการหาความคล้ายคลึงกันของผู้ใช้โดยวิเคราะห์จากลูกค้าที่มีพฤติกรรมใกล้เคียงกันด้วยสมการการหาความคล้ายโคลไซน์ ซึ่งเป็นฟังก์ชันในภาษา Python ในการพัฒนาระบบแนะนำรายนต์ให้กับลูกค้า ด้วยสมการความคล้ายโคลไซน์ (cosine similarity) จากการทดลองเมื่อนำข้อมูลมาจัดลำดับคะแนนความชอบของผู้ใช้แต่ละคน เพื่อเป็นการเพิ่มความเร็วให้อัลกอริทึมของวิธีการกรองแบบร่วมมือ อีกทั้งระบบจะนำความคล้ายคลึงโคลไซน์ของผู้ใช้ในระบบกับผู้ใช้เบ้าหมายมาทดสอบความแม่นยำของระบบด้วยค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ พบว่ามีค่าเท่ากับ 0.97 เมื่อกำหนดค่า k ไว้เท่ากับ 5 สรุปได้ว่าระบบมีประสิทธิภาพในการแนะนำรุ่นรถยนต์ที่รวดเร็วและความแม่นยำอยู่ในระดับที่ดี (Warakorn Pradiskul, 2021)

บทที่ 3 ขั้นตอนการดำเนินงาน

การจัดทำโครงการทางวิทยาการคอมพิวเตอร์ ระบบแนะนำบริษัทสำหรับผู้ประกอบการตามความสนใจเดียวๆ เทคโนโลยีปัญญาประดิษฐ์ ผู้จัดทำจึงได้กำหนดขั้นตอนการดำเนินงานดังนี้

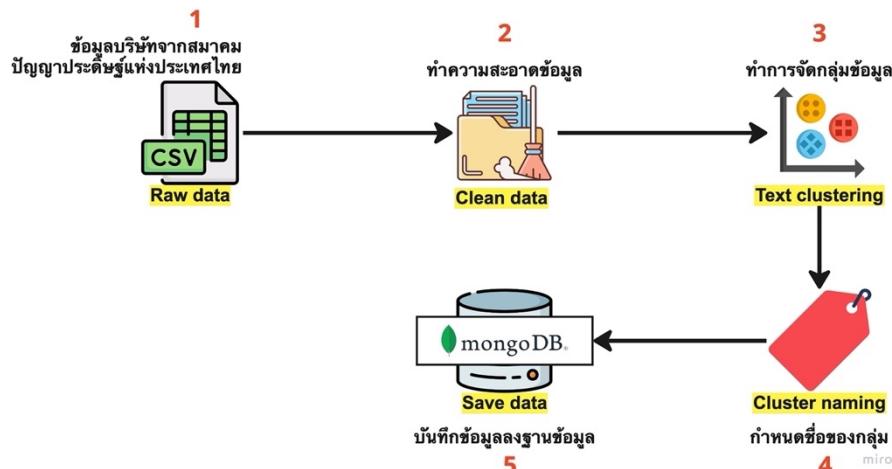
- 3.1 การเตรียมและวิเคราะห์ข้อมูล
- 3.2 การทำงานของระบบ
- 3.3 การวิเคราะห์และออกแบบระบบ
- 3.4 การออกแบบฐานข้อมูล
- 3.5 การออกแบบหน้าจอ

3.1 การเตรียมและวิเคราะห์ข้อมูล

ตาราง 4 การวิเคราะห์ข้อมูล

ข้อมูล	จำนวน	หมาย
จำนวนข้อมูลสถานประกอบการทั้งหมด	1643	รายการ
Artificial Intelligence	31	รายการ
internet of things	105	รายการ
Chatbot	46	รายการ
Big data	84	รายการ
Machine learning	61	รายการ
Data science	102	รายการ
face recognition	20	รายการ
face detection	5	รายการ
optical character recognition	3	รายการ
data mining	9	รายการ
natural language processing	11	รายการ
data visualization	1	รายการ
image processing	20	รายการ
robotics	45	รายการ
computer vision	14	รายการ
speech recognition	2	รายการ
automatic license plate recognition	1	รายการ
e-kyc	1	รายการ
biometrics	9	รายการ
biometric authentication	3	รายการ
sentiment analysis	3	รายการ
text mining	2	รายการ
embedded system	1	รายการ

machine translation	1	รายการ
ไม่มีประเภท	1318	รายการ
จำนวนคำทั้งหมด	9856	คำ

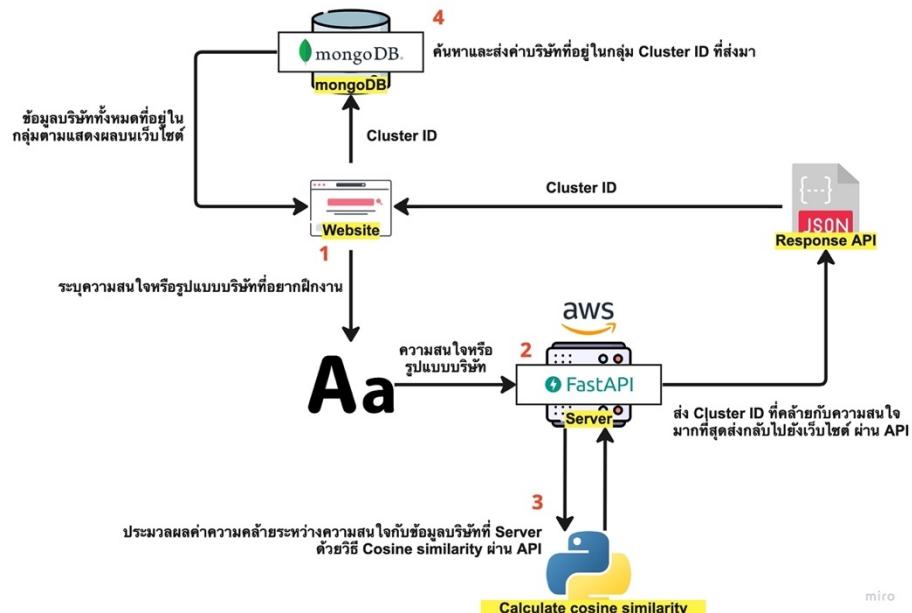


ภาพที่ 10 การเตรียมและวิเคราะห์ข้อมูล

จากภาพที่ 10 แสดงการเตรียมและวิเคราะห์ข้อมูลได้ดังนี้

- ข้อมูลบริษัทจากสมาคมนักข่าวประจำตัวชื่อแห่งประเทศไทยโดยเป็นไฟล์ข้อมูลแบบ CSV (Comma-Separated Value)
- ทำความสะอาดข้อมูลลบข้อมูลที่ไม่มีความหมายในตัว ลบตัวเลขที่ไม่จำเป็น แก้ไขพิมพ์ผิดและอักษรพิเศษต่าง ๆ
- เมื่อได้ข้อมูลที่พร้อมแล้วทำการจัดกลุ่มข้อมูลด้วยวิธี K-Means clustering
- ตั้งชื่อของกลุ่มตามข้อมูลในกลุ่มนั้น ๆ
- จัดเก็บลงฐานข้อมูล (MongoDB) หลังจากจัดกลุ่มแล้ว

3.2 การทำงานของระบบ



ภาพที่ 11 การทำงานของระบบ

จากภาพที่ 11 แสดงการทำงานของระบบโดยดังนี้

- Website ใช้ระบุความสนใจเพื่อส่งค่าไปประมวลผลความคล้ายคลึงกับข้อมูลบริษัท
- Server ใช้ประมวลผลความคล้ายคลึงกันระหว่างความสนใจที่ได้รับมาและข้อมูลบริษัทที่อยู่ในฐานข้อมูลด้วยเทคนิค Cosine similarity โดยภาษา Python และส่งค่ากลับไปเป็น Cluster ID ผ่าน API
- Response API เป็นข้อมูลในรูปแบบ JSON ที่มีค่า Cosine similarity ของแต่ละกลุ่มและ Cluster ID ที่มีความคล้ายคลึงกับความสนใจมากที่สุด
- mongoDB เป็นฐานข้อมูลที่เก็บข้อมูลบริษัทไว้ และรอเว็บไซต์เรียกใช้ข้อมูลเพื่อนำไปแสดงผล

3.3 การวิเคราะห์และออกแบบระบบ

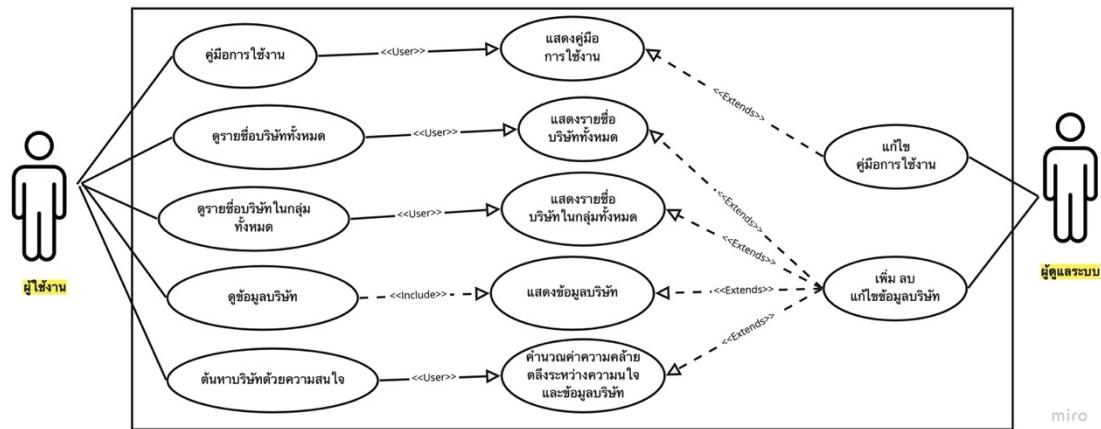
3.3.1 การวิเคราะห์ระบบ

การวิเคราะห์ระบบและการออกแบบ (System Analysis and Design) คือ วิธีการที่ใช้ในการสร้างระบบสารสนเทศขึ้นมาใหม่ในธุรกิจ ได้แก่ ห้องแม่ของธุรกิจ นอกจากการสร้างระบบสารสนเทศใหม่แล้ว การวิเคราะห์ระบบ ช่วยในการแก้ไขระบบสารสนเทศเดิมที่มีอยู่แล้ว ให้ดีขึ้น การวิเคราะห์ระบบ คือ การหาความต้องการ (Requirements) ของระบบสารสนเทศว่า คืออะไร หรือต้องการเพิ่มเติมอะไรเข้ามาในระบบ การออกแบบ คือ การนำความต้องการของระบบมาเป็นแบบแผน หรือเรียกว่า พิมพ์เขียวในการสร้างระบบสารสนเทศนั้นให้ใช้งานได้จริง

3.3.2 ผู้ใช้และออกแบบ (Use Case Diagram)

ผู้ใช้และออกแบบ (Use Case Diagram) คือแผนภาพที่ใช้แสดงปฏิสัมพันธ์ระหว่างระบบงานและลิงก์ที่อยู่ในระบบงาน และแสดงให้เห็นถึงส่วนประกอบทั้งหมด หรือ ภาพรวมของระบบ เป็นรากฐานในการเริ่มต้นการวิเคราะห์ระบบ โดยคนหาคำว่าระบบทำอะไร โดยไม่สนใจลักษณะ

การทำงานหรือเทคนิคการทำงานเปรียบเสมือน "กล่องดำ" โดย Use Case Diagram จะช่วยให้ผู้พัฒนาระบมสามารถแยกแยะกิจกรรมที่อาจจะเกิดขึ้นในระบบ เป็น Diagram พื้นฐาน ที่สามารถอธิบายสิ่งต่าง ๆ ได้โดยใช้รูปภาพที่ไม่ซับซ้อน



ภาพที่ 12 Use Case Diagram ระบบแนะนำบริษัทสำหรับผู้ใช้งานตามความสนใจ
ด้วยเทคโนโลยีปัญญาประดิษฐ์

ตาราง 5 คำอธิบาย Use case คู่มือการใช้งาน

Use case id:	1
Use case name:	คู่มือการใช้งาน
Actor:	ผู้ใช้งาน
Scenario:	การดูวิธีการใช้งานเว็บไซต์
Trigger event:	None
Brief Description:	อ่านวิธีการใช้งานเว็บไซต์
Purpose:	เพื่อใช้งานเว็บไซต์
Pre-condition:	เมื่อต้องการใช้งานเว็บไซต์
Main flow:	1. ผู้ใช้งานเปิดเว็บไซต์เข้าไปยังหน้าเกี่ยวกับ 2. อ่านวิธีการใช้งานเว็บไซต์
Alternate/Exceptional Flow:	None

ตาราง 6 คำอธิบาย Use case ดูรายชื่อบริษัททั้งหมด

Use case id:	2
Use case name:	ดูรายชื่อบริษัททั้งหมด
Actor:	ผู้ใช้งาน
Scenario:	การดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Trigger event:	None
Brief Description:	ดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Purpose:	ดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Pre-condition:	เมื่อต้องการดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล

ตาราง 5 (ต่อ)

Main flow:	1. ผู้ใช้งานเปิดเว็บไซต์เข้าไปยังหน้ารายชื่อบริษัททั้งหมด 2. ดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Alternate/Exceptional Flow:	None

ตาราง 7 คำอธิบาย Use case ดูรายชื่อบริษัทในกลุ่มทั้งหมด

Use case id:	3
Use case name:	ดูรายชื่อบริษัทในกลุ่มทั้งหมด
Actor:	ผู้ใช้งาน
Scenario:	การดูรายชื่อบริษัทในกลุ่มทั้งหมด
Trigger event:	กรณีที่แสดงผลจากการค้นหาด้วยความสนใจ หรือกรณีที่คลิกเมนูกลุ่มของบริษัท
Brief Description:	ดูรายชื่อบริษัทในกลุ่มที่ต้องการทั้งหมด
Purpose:	เพื่อดูรายชื่อบริษัทในกลุ่มทั้งหมด
Pre-condition:	เมื่อต้องการดูรายชื่อบริษัทในกลุ่มทั้งหมด
Main flow:	1. ผู้ใช้งานเปิดเว็บไซต์เข้าไปยังหน้ากลุ่มบริษัทที่ต้องการ 2. ดูรายชื่อบริษัทในกลุ่มทั้งหมด
Alternate/Exceptional Flow:	None

ตาราง 8 คำอธิบาย Use case ดูข้อมูลบริษัท

Use case id:	4
Use case name:	ดูข้อมูลบริษัท
Actor:	ผู้ใช้งาน
Scenario:	การดูข้อมูลบริษัท
Trigger event:	None
Brief Description:	ดูข้อมูลบริษัท เช่น ข้อมูลติดต่อ จังหวัด และรูปแบบธุรกิจ
Purpose:	เพื่อดูข้อมูลบริษัท
Pre-condition:	เมื่อต้องการดูข้อมูลบริษัท
Main flow:	1. ผู้ใช้งานคลิกที่เมนูชื่อของบริษัทที่ต้องการดูข้อมูล 2. ดูข้อมูลบริษัท
Alternate/Exceptional Flow:	None

ตาราง 9 คำอธิบาย Use case คนหาบริษัทด้วยความสนใจ

Use case id:	5
Use case name:	คนหาบริษัทด้วยความสนใจ
Actor:	ผู้ใช้งาน
Scenario:	การคนหาบริษัทด้วยความสนใจ
Trigger event:	None
Brief Description:	คนหาบริษัทด้วยความสนใจหรือรูปแบบธุรกิจ

ตาราง 8 (ต่อ)

Purpose:	เพื่อคนหาบริษัทด้วยความสนใจ
Pre-condition:	เมื่อต้องการคนหาบริษัทด้วยความสนใจ
Main flow:	<ol style="list-style-type: none"> ผู้ใช้งานพิมพ์ความสนใจหรือรูปแบบธุรกิจที่ช่องคนหาและ แสดงรายชื่อบริษัทที่อยู่ในกลุ่มที่ระบบแนะนำ ผู้ใช้คลิกเลือกบริษัทเพื่อดูข้อมูลบริษัท
Alternate/Exceptional Flow:	None

ตาราง 10 คำอธิบาย Use case แก้ไขคู่มือการใช้งาน

Use case id:	6
Use case name:	แก้ไขคู่มือการใช้งาน
Actor:	ผู้ดูแลระบบ
Scenario:	การแก้ไขคู่มือการใช้งาน
Trigger event:	None
Brief Description:	แก้ไขคู่มือการใช้งาน
Purpose:	เพื่อแก้ไข คู่มือการใช้งาน
Pre-condition:	เมื่อต้องการแก้ไขคู่มือการใช้งาน
Main flow:	<ol style="list-style-type: none"> ผู้ดูแลระบบแก้ไขข้อมูลคู่มือการใช้งาน Deploy เพื่ออัปเดตระบบ
Alternate/Exceptional Flow:	None

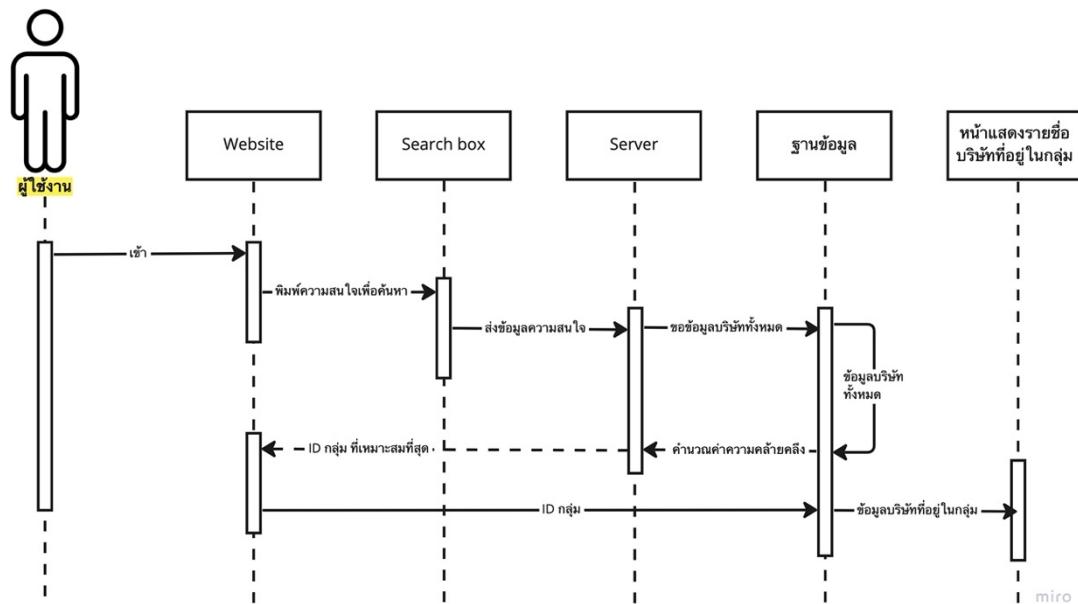
ตาราง 11 คำอธิบาย Use case เพิ่ม ลบ แก้ไขข้อมูลบริษัท

Use case id:	7
Use case name:	เพิ่ม ลบ แก้ไขข้อมูลบริษัท
Actor:	ผู้ดูแลระบบ
Scenario:	การเพิ่ม ลบ แก้ไขข้อมูลบริษัท
Trigger event:	None
Brief Description:	เพิ่ม ลบ แก้ไขข้อมูลบริษัท
Purpose:	เพื่อเพิ่ม ลบ แก้ไขข้อมูลบริษัท
Pre-condition:	เมื่อต้องการเพิ่ม ลบ แก้ไขข้อมูลบริษัท
Main flow:	1. ผู้ดูแลระบบเพิ่ม ลบ แก้ไขข้อมูลบริษัท ในฐานข้อมูล
Alternate/Exceptional Flow:	None

3.2.3 ชีวนิรนัยโดยแกรม (Sequence Diagram)

Sequence Diagram เป็นหนึ่งในแผนผังการทำงานแบบ Unified Modeling Language (UML) ใช้สำหรับการสร้างแบบจำลองเชิงวัตถุ โดยขอแตกต่างจากแผนผังรูปแบบ UML อื่น คือเป็น แผนผังการทำงานที่แสดงลำดับการปฏิสัมพันธ์ (Sequence of interactions) ระหว่างวัตถุที่แสดงภายในระบบต่างๆ อาทิ เช่น การส่งข้อความ (messaging) ที่มีการรับส่งข้อมูลระหว่างผู้ใช้ Sequence Diagram เป็นแผนผังการทำงานที่ประกอบไปด้วยคลาส (Class) หรือวัตถุ (Object) เส้นประที่ใช้เพื่อ

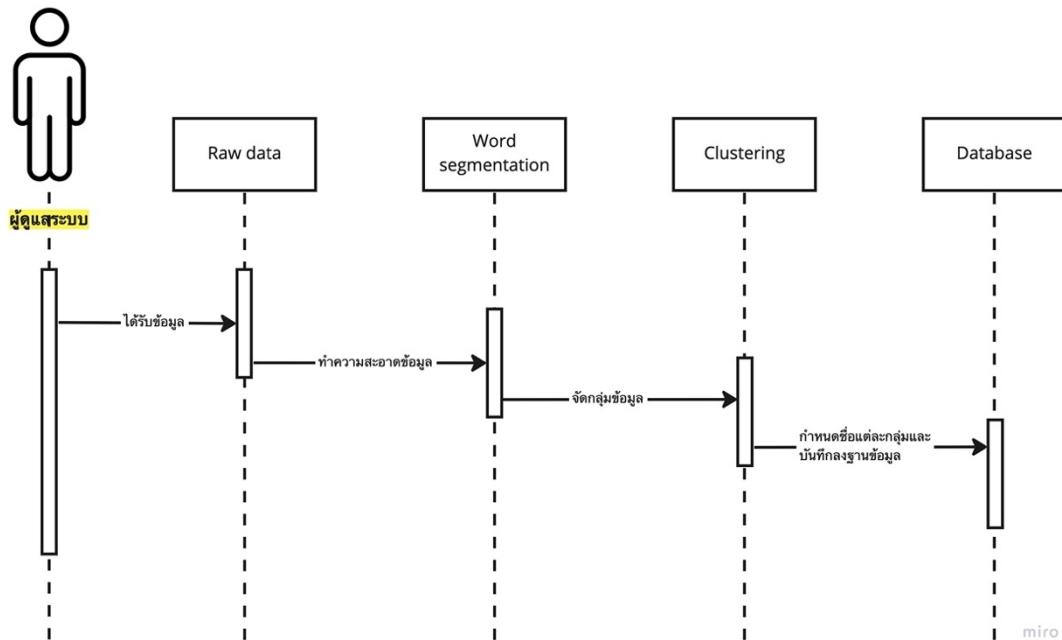
แสดงลำดับเวลา และเส้นที่ใช้เพื่อแสดงกิจกรรมที่เกิดขึ้นจากคลาสรหีอวัตถุในแผนผังการทำงานภายใน Sequence Diagram จะใช้สีเหลี่ยมแทนสเมือนคลาสและวัตถุโดยภายในจะมีชื่อของคลาสหรือวัตถุประกอบอยู่ในรูปแบบ [Object]: Class



ภาพที่ 13 Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้

ตาราง 12 อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้

เหตุการณ์ที่เกิดขึ้น	คำอธิบาย
เข้า Website	เข้า Website ด้วย Browser
พิมพ์ความสนใจเพื่อค้นหา	ระบุความสนใจรูปแบบคร่าวกิจของบริษัทหรือความสนใจที่อยากฝึกงานของผู้ใช้
ส่งข้อมูลความสนใจ	ส่งข้อมูลความสนใจที่ผู้ใช้ระบุไปประมวลผลที่ Server
ขอข้อมูลบริษัททั้งหมด	Server ขอข้อมูลบริษัททั้งหมดจากฐานข้อมูลเพื่อนำมาเก็บไว้รอคำนวนความคล้ายคลึง
ข้อมูลบริษัททั้งหมด	ข้อมูลบริษัททั้งหมดในฐานข้อมูล ส่งให้ Server
คำนวนความคล้ายคลึง	คำนวนความคล้ายคลึงระหว่างความสนใจของผู้ใช้และข้อมูลบริษัท
ID กลุ่มที่เหมาะสมที่สุด	คืนค่า ID ของกลุ่มบริษัทที่คล้ายกับความสนใจของผู้ใช้
ID กลุ่ม	ส่งค่า ID ของกลุ่มไปยังฐานข้อมูลเพื่อขอข้อมูลบริษัทที่อยู่ในกลุ่มนั้น ๆ
ข้อมูลบริษัทที่อยู่ในกลุ่ม	แสดงรายชื่อบริษัทที่อยู่ในกลุ่มในหน้าเว็บ



ภาพที่ 14 Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบัญชีใหม่

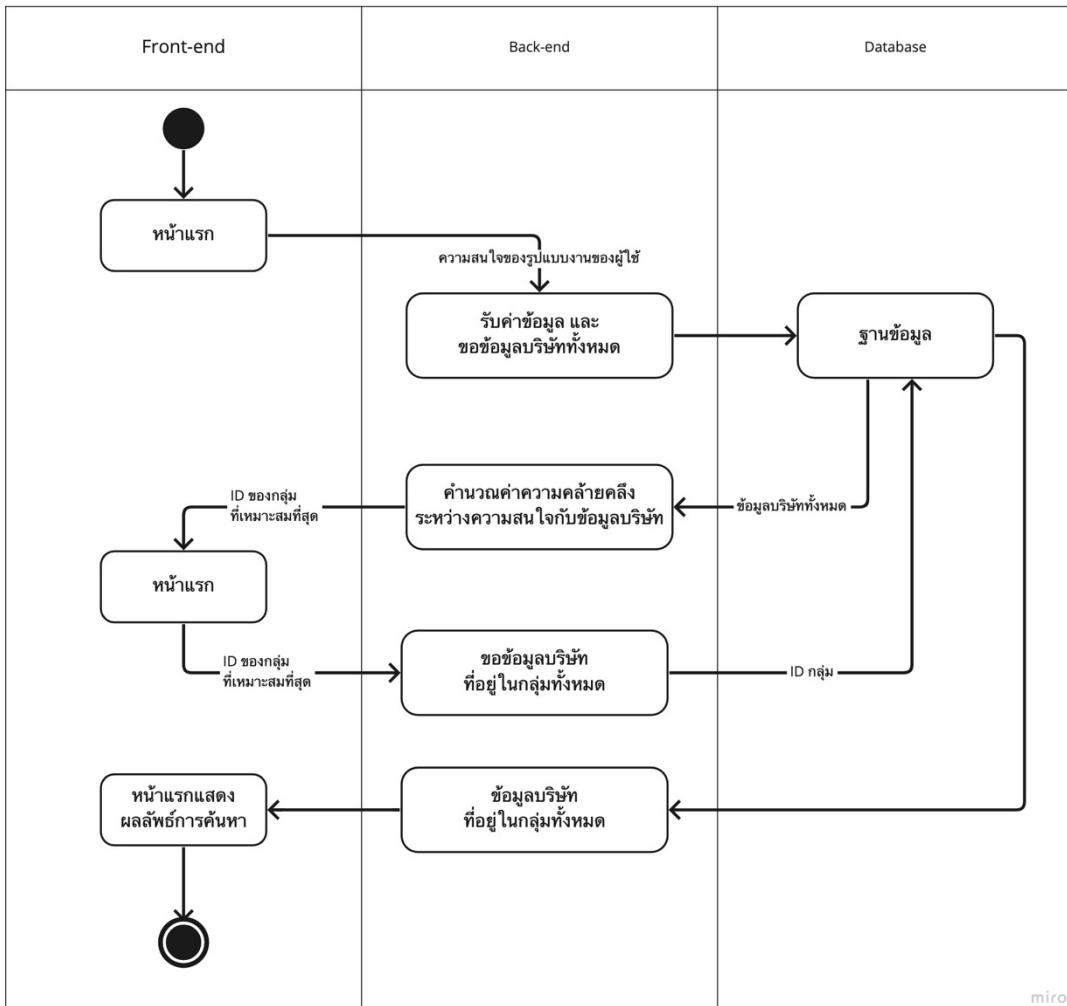
ตาราง 13 อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบัญชีใหม่

เหตุการณ์ที่เกิดขึ้น	คำอธิบาย
ได้รับข้อมูล	ได้รับข้อมูลดิบที่จะนำมาใช้งาน
ทำการตรวจสอบข้อมูล	นำข้อมูลดิบมาทำการลบตัวเลข คำที่ไม่มีความหมายในตัว คำສະกัดผิด
จัดกลุ่มข้อมูล	ทำการหาคำสำคัญและทำการจัดกลุ่มข้อมูล
กำหนดชื่อแต่ละกลุ่มและบันทึกลงฐานข้อมูล	กำหนดชื่อของกลุ่มและบันทึกข้อมูลลงฐานข้อมูลเพื่อใช้ในเว็บไซต์

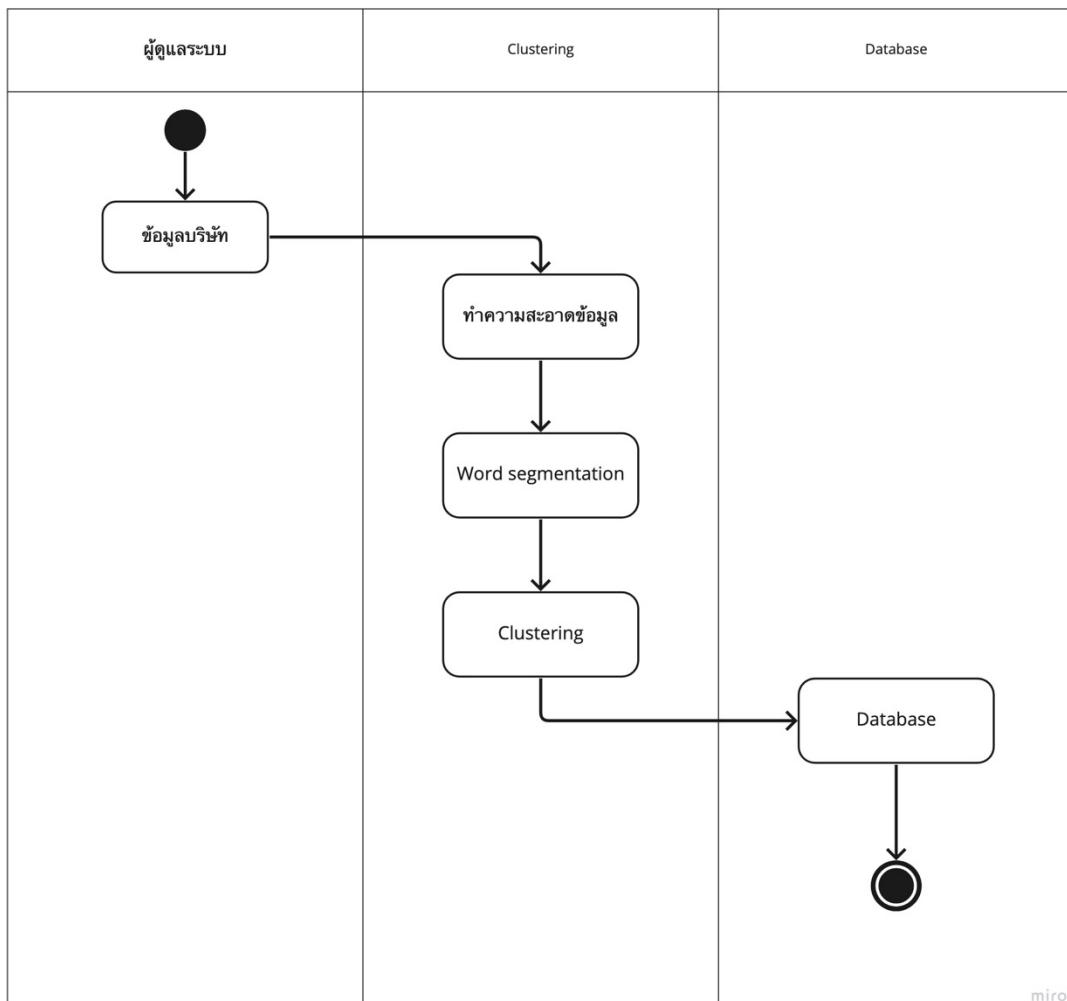
3.3.4 แอคทิวิตี้ไดอะแกรม (Activity Diagram)

Activity Diagram หรือแผนภาพกิจกรรม ใช้อธิบายกิจกรรมที่เกิดขึ้นในลักษณะกราฟแสดงการให้ผลของการทำงาน (Workflow) จะมีลักษณะเดียวกับ Flowchart โดยขั้นตอนในการทำงานแต่ละขั้นจะเรียกว่า Activity ดังภาพที่ 16 และภาพที่ 17 มีรายละเอียดดังนี้ การใช้งาน Activity Diagram

1. อธิบายกราฟแสดงการให้ผลของการทำงาน (Workflow)
2. แสดงขั้นตอนการทำงานของระบบ



ภาพที่ 15 Activity Diagram ของผู้ใช้งาน



ภาพที่ 16 Activity Diagram ของผู้ดูแลระบบ

3.4 การออกแบบฐานข้อมูล

3.4.1 ER Diagram

ER Diagram คือ แบบจำลองที่ใช้เขียนโดยโครงสร้างของฐานข้อมูลซึ่งเขียนออกแบบมาในลักษณะของรูปภาพ การเขียนโดยโครงสร้างและความสัมพันธ์ของข้อมูล (Relationship) ประกอบด้วย 1.เอนทิตี้ (Entity) เป็นวัตถุ หรือสิ่งของที่เราสนใจในระบบงานนั้น 2.แอทริบิวต์ (Attribute) เป็นคุณสมบัติของวัตถุที่เราสนใจ 3.ความสัมพันธ์ (Relationship) คือ ความสัมพันธ์ระหว่างเอนทิตี้ ER Diagram มีความสำคัญต่อการพัฒนาระบบงานฐานข้อมูล Application ต่างๆ ที่ต้องการการเก็บข้อมูลอย่างมีระบบ มีโครงสร้าง ดังนั้น ER Diagram จึงใช้เพื่อเป็นเอกสารในการสื่อสารระหว่างนักออกแบบระบบ และนักพัฒนาระบบ เพื่อให้สื่อสารอย่างตรงกัน

companies	
_id	PK
short_company	
th_company_name	
eng_company_name	
type_business	
product	
type_innovation	
detail	
owner	
province_base	
address	
phone_number	
email	
website	
source	
cluster	

ภาพที่ 17 ER Diagram ระบบแนะนำบริษัทสำหรับผู้ใช้งานตามความสนใจ
ด้วยเทคโนโลยีปัญญาประดิษฐ์

3.4.2 พจนานุกรมข้อมูล (Data Dictionary)

หลังจากที่วิเคราะห์ระบบแล้ว ผู้ศึกษาได้ออกแบบฐานข้อมูล โดยออกแบบ
โครงสร้างของระบบ ซึ่งประกอบไปด้วยตาราง จำนวน 1 ตาราง และได้อธิบาย ชื่อ ตาราง(File
Name), คำอธิบาย(Description), ชื่อข้อมูล(Field Name), ชนิดของข้อมูล(Type), ขนาดที่เก็บ(Length),
ลักษณะที่เก็บค่า(Format), ชนิดของคีย์(Key) ดังต่อไปนี้

ตาราง 14 พจนานุกรมข้อมูลบริษัท

File name: companies					
Description: ตารางเก็บข้อมูลบริษัททั้งหมด					
Field name	Type	Length	Format	Description	Key
_id	String	50	ตัวอักษร	รหัสบริษัท	Primary key
short_company	String	50	ตัวอักษร	ชื่อย่อบริษัท	Null
th_company_name	String	50	ตัวอักษร	ชื่อบริษัท ภาษาไทย	Null
eng_company_name	String	50	ตัวอักษร	ชื่อบริษัท ภาษาอังกฤษ	Null
type_business	String	50	ตัวอักษร	ประเภท ธุรกิจ	Null
product	String	255	ตัวอักษร	ประเภท สินค้า	Null
type_innovation	String	50	ตัวอักษร	ประเภท เทคโนโลยี	Null
detail	String	255	ตัวอักษร	รายละเอียด ธุรกิจ	Null
owner	String	255	ตัวอักษร	เจ้าของ	Null

ตาราง 13 (ต่อ)

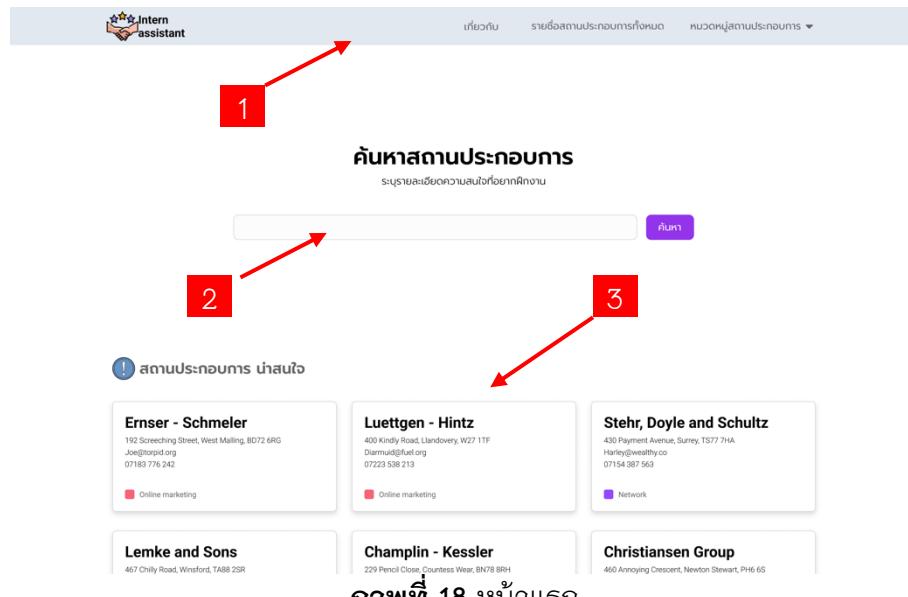
province_base	String	50	ตัวอักษร	จังหวัดที่ตั้ง	Null
address	String	255	ตัวอักษร	ที่อยู่โดยละเอียด	Null
phone_number	String	20	ตัวอักษร	เบอร์โทร	Null
email	String	50	ตัวอักษร	อีเมล	Null
website	String	50	ตัวอักษร	เว็บไซต์	Null
source	String	255	ตัวอักษร	ที่มาของมูล	Null
cluster	String	1	ตัวอักษร	กลุ่ม	Not Null

3.5 การออกแบบหน้าจอ

การออกแบบหน้าจอหรือ UI design นั้นเป็นส่วนที่ผู้พัฒนาโปรแกรมต้องทำเนื่องจากหน้าจอ นั้นเป็นส่วนสำคัญที่จะเชื่อมต่อกับผู้ใช้งานโปรแกรมหรือเก็คิลส่วนที่ผู้ใช้งานจะเห็น สั่งการ และติดตามได้ เช่น สามารถออกแบบในทางด้านของหน้าตา ปุ่ม ช่องที่ใช้พิมพ์สำหรับคนหา ขนาดตัวอักษร สี และรูปภาพเป็นต้น

การออกแบบหน้าจอสำหรับการค้นหาสถานประกอบการสำหรับผู้ใช้งานด้วยเทคโนโลยีปัญญาประดิษฐ์นั้น จะเน้นในแพลตฟอร์มที่เป็นหน้าจอคอมพิวเตอร์หรือ Desktop เป็นหลัก โดยมีรายละเอียดดังนี้

3.5.1 หน้าแรก



ภาพที่ 18 หน้าแรก

หมายเลขอ 1 เมนูหลัก

หมายเลขอ 2 ช่องค้นหา

หมายเลขอ 3 รายชื่อบริษัทแนะนำ

3.5.2 หน้าเกี่ยวกับ

The screenshot shows a user interface for managing data. At the top, there is a navigation bar with icons for 'Intern assistant', 'เกี่ยวกับ' (About), 'รายชื่อสถานประกอบการทั้งหมด' (List of all business operators), and 'หน้าดูผู้สถานประกอบการ' (Business operator view). A red arrow labeled '1' points to the 'เกี่ยวกับ' button. Below the navigation, there is a text area with placeholder text. A red arrow labeled '2' points to a diagram illustrating a data flow: a CSV file is imported into a central processing area, which then outputs to a database.

ภาพที่ 19 หน้าเกี่ยวกับ

หมายเลขอ 1 เมนูหลัก

หมายเลขอ 2 รายละเอียดเกี่ยวกับเว็บไซต์

3.5.3 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

The screenshot shows a search results page for 'Online marketing'. At the top, there is a navigation bar with icons for 'Intern assistant', 'เกี่ยวกับ' (About), 'รายชื่อสถานประกอบการทั้งหมด' (List of all business operators), and 'หน้าดูผู้สถานประกอบการ' (Business operator view). A red arrow labeled '1' points to the search input field. The search results list various company names, each with a small icon and a 'View' button. A red arrow labeled '2' points to the 'Olson LLC' entry in the list.

Ernsler - Schmeler	Von, Kunde And Stracke
Tromp LLC	Bode And Sons
Gusikowski - Considine	Grady, Huels And Runte
Dicki, Welch And Rippin	Balistreri Inc
Boyer - Daugherty	Cartwright, Goyette And Watsica
Gutkowski - Kautzer	Schulist LLC
Ledner - Pfeffer	Olson LLC
Mohr, Herzog And Terry	Davis, Dibbert And Schuster
Harris, Veum And Kertzmann	Ferry - Howe
Rolfson - Wiza	Konopelski Group
Baumbach - Raynor	Rutherford - Denesik
Dach, Cummings And Lindgren	Grant And Sons
Halvorson - Nikolaus	Wilkinson - Hegmann
Mueller - Hodkiewicz	Conn - Bergstrom

ภาพที่ 20 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หมายเลขอ 1 เมนูหลัก

หมายเลขอ 2 รายชื่อบริษัททั้งหมดที่อยู่ในกลุ่ม

3.5.4 หน้าแสดงรายชื่อบริษัททั้งหมด

1

2

สถานะประกอบการทั้งหมด	
Ernsler - Schmeler	ก่อตั้งมาแล้ว
Tromp LLC	ก่อตั้งมาแล้ว
Gusikowski - Considine	ก่อตั้งมาแล้ว
Dicki, Welch And Rippin	บ้าน
Boyer - Daugherty	สำนักงาน
Gutkowski - Kautzer	เมืองใหญ่
Ledner - Pfeffer	เมืองใหญ่
Mohr, Herzog And Terry	เมืองใหญ่
Harris, Veum And Kertzmann	เมืองใหญ่
Rolfson - Wiza	เมืองใหญ่
Baumbach - Raynor	เมืองใหญ่
Dach, Cummings And Lindgren	เมืองใหญ่
Halvorson - Nikolaus	เมืองใหญ่
Mueller - Hodkiewicz	เมืองใหญ่
Von, Kunde And Stracke	ก่อตั้งมาแล้ว
Bode And Sons	ก่อตั้งมาแล้ว
Grady, Huels And Runte	ก่อตั้งมาแล้ว
Ballistreri Inc	สำนักงาน
Cartwright, Goyette And Watsica	สำนักงาน
Schulist LLC	เมืองใหญ่
Olson LLC	เมืองใหญ่
Davis, Dibbert And Schuster	เมืองใหญ่
Ferry - Howe	เมืองใหญ่
Konopelski Group	เมืองใหญ่
Rutherford - Denesik	เมืองใหญ่
Grant And Sons	เมืองใหญ่
Wilkinson - Hegmann	เมืองใหญ่
Conn - Bergstrom	เมืองใหญ่

ภาพที่ 21 หน้าแสดงรายชื่อบริษัททั้งหมด

หมายเลขอ 1 เมนูหลัก

หมายเลขอ 2 รายชื่อบริษัททั้งหมด

3.5.5 หน้าแสดงผลลัพธ์รายชื่อบริษัท

1

2

3

ค้นหาสถานประกอบการ	
ระบุรายละเอียดความสนใจที่อยากศึกษา	
หัวข้อ SEO	x
ค้นหา	
Ernsler - Schmeler	ก่อตั้งมาแล้ว
Tromp LLC	ก่อตั้งมาแล้ว
Gusikowski - Considine	ก่อตั้งมาแล้ว
Dicki, Welch And Rippin	บ้าน
Boyer - Daugherty	สำนักงาน
Von, Kunde And Stracke	ก่อตั้งมาแล้ว
Bode And Sons	ก่อตั้งมาแล้ว
Grady, Huels And Runte	ก่อตั้งมาแล้ว
Ballistreri Inc	สำนักงาน
Cartwright, Goyette And Watsica	สำนักงาน

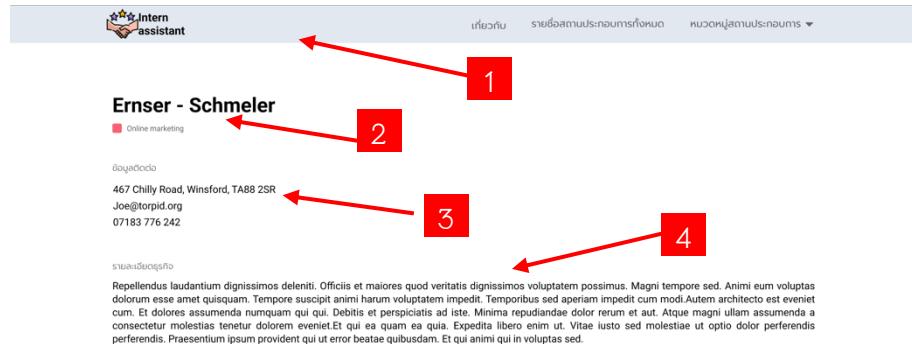
ภาพที่ 22 หน้าแสดงผลลัพธ์รายชื่อบริษัท

หมายเลขอ 1 เมนูหลัก

หมายเลขอ 2 ซองคนหา

หมายเลขอ 3 ผลลัพธ์รายชื่อบริษัทที่อยู่ในกลุ่มที่คล้ายคลึงกับความสนใจ

๓.๕.๖ หน้าแสดงข้อมูลบริษัท



ภาพที่ 23 หน้าแสดงข้อมูลบริษัท

- หมายเลขอ 1 เมนูหลัก
- หมายเลขอ 2 ชื่อบริษัท
- หมายเลขอ 3 ที่อยู่และข้อมูลติดต่อของบริษัท
- หมายเลขอ 4 รูปแบบรายละเอียดธุรกิจของบริษัท



บทที่ 4

ผลการดำเนินงาน

การทำงานของระบบแนะนำนำบริษัทสำหรับผู้คนตามความสนใจด้วยเทคโนโลยีปัญญาประดิษฐ์ มีขั้นตอนการทำงานดังนี้

4.1 การวิเคราะห์และการทำ Word segmentation

4.1.1 การวัดค่าความแม่นยำในการตัดคำ

4.1.2 การจัดกลุ่มข้อมูล

4.1.3 การกำหนดชื่อกลุ่ม

4.1.1 การวัดค่าความแม่นยำในการตัดคำ

ในขั้นตอนการทำ Word segmentation นั้นมีขั้นตอนอยู่อย่างในการทำร่วมด้วยหลายขั้นตอนหลังจากทำความสะอาดข้อมูล คือการทำคำสำคัญของแต่ละประโยคในที่นี่คือรายละเอียดธุรกิจของแต่ละบริษัท การกำจัดคำที่ไม่สืบความหมายหรือ Stop word เช่น ที่ นี่ อีน เป็นต้น ออกไปจากประโยคเพื่อให้ได้ประโยคที่มีเนื้อหาใจความดีที่สุด และอีกขั้นตอนสำคัญคือการทำคำศัพท์ เว็บดของแต่ละประโยคเพื่อที่จะได้ทราบว่าประโยคนั้น ๆ กำลังสื่อถึงเรื่องไหนเป็นสำคัญด้วยเทคนิค TF-IDF ที่เป็นการทำหน้าหนักของคำนั้น ๆ ในประโยค

ในการทำงานประมวลผลเกี่ยวกับการประมวลผลภาษาธรรมชาติจำเป็นต้องมีการตัดคำออกเป็นคำ ๆ เพื่อจะได้่ายและนำไปเข้าสู่กระบวนการต่าง ๆ ได้อย่างง่ายโดยปกติแล้วการตัดคำในภาษาอังกฤษนั้นสามารถตัดได้โดยใช้การเว้นวรรคเป็นเงื่อนไขในการตัด แต่ในภาษาไทยนั้นการเขียนนั้นไม่ได้มีการเว้นวรรคคำเหมือนภาษาอังกฤษทำให้การตัดคำคำนั้นจะใช้เว้นวรรคมาตัดคำตลอดไม่ได้ จำเป็นต้องใช้อัลกอริทึมอื่น ๆ เช่นมาช่วย เช่นการใช้ Dictionary-based, Maximum Matching เป็นต้น

และในภาษา Python ก็มีเครื่องมือที่สามารถช่วยอำนวยความสะดวกในการตัดคำภาษาไทยอย่าง Pythainlp ที่ผู้วิจัยได้เลือกใช้ในโครงการนี้ในไลบรารีนั้นสามารถตัดคำได้หลาย Engine ด้วยกันและในแต่ละตัวเลือกที่ใช้อัลกอริทึมต่างกันยกตัวอย่างเช่น newmm, longest, newmm-safe, mm, icu, deepcut, attacut เป็นต้น

1. การวัดค่าความแม่นยำในการตัดคำ

ก่อนที่จะเลือก Engine มาใช้ตัดคำจำเป็นต้องมีการวัดค่าความถูกต้องเพื่อที่จะได้ผลลัพธ์ที่ตรงกับความต้องการมากที่สุดและในโครงการนี้ผู้วิจัยได้เลือก Engine มาทดสอบด้วยกันจำนวน 3 ตัวเลือกดังนี้

1. newmm – dictionary-based, Maximum Matching + Thai Character Cluster

2. deepcut – wrapper for DeepCut, learning-based approach

3. longest – dictionary-based, Longest Matching

วิธีที่ทดสอบคือทำการสุมเลือกประโยคมาจำนวน 100 ประโยค ทำการตัดคำในแต่ละประโยคเองโดยไม่ใช้ตัวช่วย ใช้ไลบรารี Pythainlp ในแต่ละ Engine ตัดคำและนำมาเทียบกับประโยคที่ผู้วิจัยตัดไว้โดย Engine ใหม่ที่มีความเหมือนกับที่ผู้วิจัยตัดไว้มากที่สุดก็จะถือว่ามีความใกล้เคียงกับความต้องการของผู้วิจัยมากที่สุด

longest_df.csv

```
[{"บริการ": "platform", "ชื่อ": "แอปพลิเคชัน", "คำอธิบาย": "สำหรับสื่อสารทางดิจิตัล", "เทคโนโลยี": "HTML, CSS, JavaScript", "ตัวอย่าง": "Facebook, Instagram"}, {"บริการ": "platform", "ชื่อ": "แพลตฟอร์ม", "คำอธิบาย": "เครื่องมือในการทำงาน", "เทคโนโลยี": "Python, PostgreSQL", "ตัวอย่าง": "Dropbox, Google Sheets"}, {"บริการ": "platform", "ชื่อ": "เว็บไซต์", "คำอธิบาย": "แพลตฟอร์มสำหรับการค้าอิเล็กทรอนิกส์", "เทคโนโลยี": "Java, Spring Boot", "ตัวอย่าง": "Amazon, eBay"}, {"บริการ": "platform", "ชื่อ": "ซอฟต์แวร์", "คำอธิบาย": "ซอฟต์แวร์สำหรับธุรกิจ", "เทคโนโลยี": "C++, Python", "ตัวอย่าง": "SAP ERP, Microsoft Dynamics"}, {"บริการ": "platform", "ชื่อ": "โซลูชัน", "คำอธิบาย": "โซลูชันที่เฉพาะเจาะจง", "เทคโนโลยี": "Angular, React", "ตัวอย่าง": "Oracle Database, Oracle Database"}, {"บริการ": "platform", "ชื่อ": "เว็บแอป", "คำอธิบาย": "แอปพลิเคชันที่ใช้ในเบราว์เซอร์", "เทคโนโลยี": "Node.js, Express.js", "ตัวอย่าง": "Netflix, Netflix"}, {"บริการ": "platform", "ชื่อ": "เดสก์ท็อป", "คำอธิบาย": "แอปพลิเคชันที่ต้องติดตั้งบนคอมพิวเตอร์", "เทคโนโลยี": "Java Swing, JavaFX", "ตัวอย่าง": "Adobe Photoshop, Microsoft Word"}, {"บริการ": "platform", "ชื่อ": "เว็บไซต์", "คำอธิบาย": "เว็บไซต์ที่ต้องติดตั้งบนคอมพิวเตอร์", "เทคโนโลยี": "Windows, Mac OS", "ตัวอย่าง": "Windows 10, Mac OS X"}, {"บริการ": "platform", "ชื่อ": "ซอฟต์แวร์", "คำอธิบาย": "ซอฟต์แวร์ที่ต้องติดตั้งบนคอมพิวเตอร์", "เทคโนโลยี": "Java, C++, Python", "ตัวอย่าง": "Microsoft Office, Adobe Photoshop"}]
```

ภาพที่ 26 ตัวอย่างการตัดคำโดยใช้ Engine longest

deepcut_df.csv

```
[{"บริการ": "platform", "ชื่อ": "แอปพลิเคชัน", "คำอธิบาย": "สำหรับสื่อสารทางดิจิตัล", "เทคโนโลยี": "HTML, CSS, JavaScript", "ตัวอย่าง": "Facebook, Instagram"}, {"บริการ": "platform", "ชื่อ": "แพลตฟอร์ม", "คำอธิบาย": "เครื่องมือในการทำงาน", "เทคโนโลยี": "Python, PostgreSQL", "ตัวอย่าง": "Dropbox, Google Sheets"}, {"บริการ": "platform", "ชื่อ": "เว็บไซต์", "คำอธิบาย": "แพลตฟอร์มสำหรับการค้าอิเล็กทรอนิกส์", "เทคโนโลยี": "Java, Spring Boot", "ตัวอย่าง": "Amazon, eBay"}, {"บริการ": "platform", "ชื่อ": "ซอฟต์แวร์", "คำอธิบาย": "ซอฟต์แวร์สำหรับธุรกิจ", "เทคโนโลยี": "C++, Python", "ตัวอย่าง": "SAP ERP, Microsoft Dynamics"}, {"บริการ": "platform", "ชื่อ": "โซลูชัน", "คำอธิบาย": "โซลูชันที่เฉพาะเจาะจง", "เทคโนโลยี": "Angular, React", "ตัวอย่าง": "Oracle Database, Oracle Database"}, {"บริการ": "platform", "ชื่อ": "เว็บแอป", "คำอธิบาย": "แอปพลิเคชันที่ใช้ในเบราว์เซอร์", "เทคโนโลยี": "Node.js, Express.js", "ตัวอย่าง": "Netflix, Netflix"}, {"บริการ": "platform", "ชื่อ": "เดสก์ท็อป", "คำอธิบาย": "แอปพลิเคชันที่ต้องติดตั้งบนคอมพิวเตอร์", "เทคโนโลยี": "Java Swing, JavaFX", "ตัวอย่าง": "Adobe Photoshop, Microsoft Word"}, {"บริการ": "platform", "ชื่อ": "เว็บไซต์", "คำอธิบาย": "เว็บไซต์ที่ต้องติดตั้งบนคอมพิวเตอร์", "เทคโนโลยี": "Windows, Mac OS", "ตัวอย่าง": "Windows 10, Mac OS X"}, {"บริการ": "platform", "ชื่อ": "ซอฟต์แวร์", "คำอธิบาย": "ซอฟต์แวร์ที่ต้องติดตั้งบนคอมพิวเตอร์", "เทคโนโลยี": "Java, C++, Python", "ตัวอย่าง": "Microsoft Office, Adobe Photoshop"}]
```

ภาพที่ 27 ตัวอย่างการตัดคำโดยใช้ Engine deepcut

ผลการทดสอบความแม่นยำการตัดคำ

Engine	Accuracy (%)
newmm	90.99%
longest	83.04%
deepcut	76.65%

จากการทดสอบพบว่า Engine newmm ในไลบรารี Pythainlp นี้ มีความแม่นยำ กับที่ผู้วิจัยตัดมากที่สุดที่ 90.99% รองลงมาที่ longest ค่าความแม่นยำอยู่ที่ 83.04% และ ความแม่นยำน้อยที่สุดคือ deepcut ที่ความแม่นยำ 76.65% ตั้งภาพที่ 29 ตั้งนั้นโครงการนี้จึงใช้ Engine newmm ในการตัดคำเพื่อนำไปประมวลผลต่อในขั้นตอนอื่น ๆ

```
onze@Tinngrits-MacBook-Pro:~/documents/final_project
python calculate_accuracy.py
newmm: 90.99%
deepcut: 76.65%
longest: 83.04%
```

ภาพที่ 28 ผลการวัดค่าความแม่นยำในการตัดคำของ Engine ในไลบรารี Pythainlp

4.1.2 การจัดกลุ่มข้อมูล

เมื่อได้เครื่องมือที่จะช่วยตัดคำแล้วขั้นตอนต่อไปคือการนำมาทำการจัดกลุ่มข้อมูล โดยเทคนิคที่เลือกใช้คือการคำ TF-IDF (Term Frequency–Inverse Document Frequency) เพื่อหาว่า คำไหนในประโยคนั้นเป็นคำสำคัญของประโยคนั้น ๆ โดยการวัดจากน้ำหนักของคำด้วยวิธีดังกล่าว ตั้งภาพที่ 30 ที่แสดงการตัดคำและเรียงจากน้ำหนักของคำ 5 อันดับแรก

```
onze@Tinngrits-MacBook-Pro:~/documents/final_project
python main.py
[('based', 'datacenter', 'solutions', 'map', 'embedded')]
list(['printed', 'พิมพ์', 'พิมพ์', 'น้ำหนัก'])
list(['ขออภัย', 'ขอโทษ', 'ใจหาย', 'ขออภัย', 'ต้องเข้ามาพยายาม'])
list(['as', 'a', 'service', 'open', 'api'])
list(['ไฟฟ้า', 'ออนไลน์', 'อาชญากรรม', 'คอมเบอร์', 'ล็อก'])
list(['เทคโนโลยีสารสนเทศ', 'ธุรกิจ', 'กระบวนการ', 'ออกแบบ', 'ระบบ'])
list(['เอกสาร', 'เอกสาร', 'เอกสาร', 'เอกสาร', 'เอกสาร'])
list(['blockchain', 'technology', 'เทคโนโลยีสารสนเทศ', 'ธุรกิจ', 'ให้ค่าปรึกษา'])
list(['app', 'dolphin', 'finpoint', 'know', 'ผู้คน'])
list(['renthub', 'อพาร์ทเม้นท์', 'พื้นที่', 'เอกสาร', 'th'])
```

ภาพที่ 29 ตัวอย่างการตัดคำและเรียงคำที่มีน้ำหนักมากที่สุด 5 อันดับ

	aaa	ab	abap	abeam	ablerex	abroad	academic	acceptance	access	accessories	...	isoที่	ไอยค์ครึ่ง	isoเดียว	iso mapped	iso โนบายน	iso iso	iso ไทย	iso iso core en
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
...	
1640	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1641	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1642	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1643	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1644	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

ภาพที่ 30 ตาราง TF-IDF แสดงน้ำหนักของคำ

เมื่อได้ชุดของประโยชน์ที่ทำการตัดคำเรียบร้อยแล้วจึงนำเข้าสู่กระบวนการลบคำที่เป็น Stop word หรือคำที่ไม่สืบความหมายออกและตัวอักษรพิเศษต่าง ๆ ด้วยฟังก์ชันในไลบรารี Pythainlp และ nltk ดังภาพที่ 31 จะเห็นได้ว่ามีคำที่ถูกลบออกไปนั่นคือคำที่เป็น Stop word ยกตัวอย่างเช่น

ประโยชน์ต้น: บริการ Platform รวมสถานที่จัดประชุมและผู้ให้บริการล่าพรมงานอีเวนท์ที่ดีที่สุด
คำที่ไม่สืบความหมาย (Stop word): รวมจัดและที่ที่สุด

ประโยชน์ใหม่: บริการPlatform/สถานที่ประชุมผู้ให้บริการสำหรับงานอีเวนท์ดี

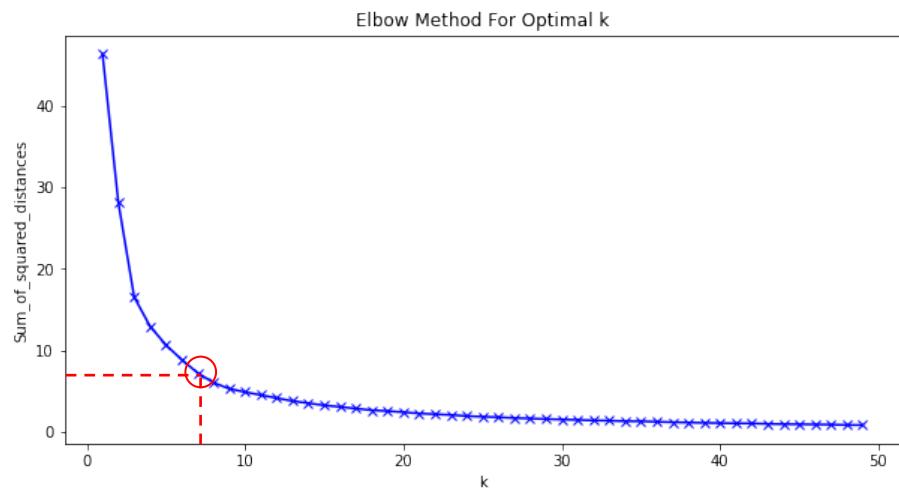
```

onze@Tinngrits-MacBook-Pro:~/documents/final_project/report/demo
python main.py
Default: บริการ Platform รวมสถานที่จัดประชุมและผู้ให้บริการล่าพรมงานอีเวนท์ที่ดีที่สุด
Keep stop word: ['บริการ', 'Platform', 'รวม', 'สถานที่', 'จัด', 'ประชุม', 'และ', 'ผู้ให้บริการ', 'ล่าพรมงาน', 'งาน', 'อีเวนท์', 'ที่', 'ดี', 'ที่สุด']
Remove stop word: ['บริการ', 'Platform', 'สถานที่', 'จัดประชุม', 'ผู้ให้บริการ', 'ล่าพรมงาน', 'งาน', 'อีเวนท์', 'ที่']
~/Documents/final_project/report !? 2> report 1>
4s backup_finalproject 18:11:01

```

ภาพที่ 31 ตัวอย่างการตัดคำและลบ Stop word

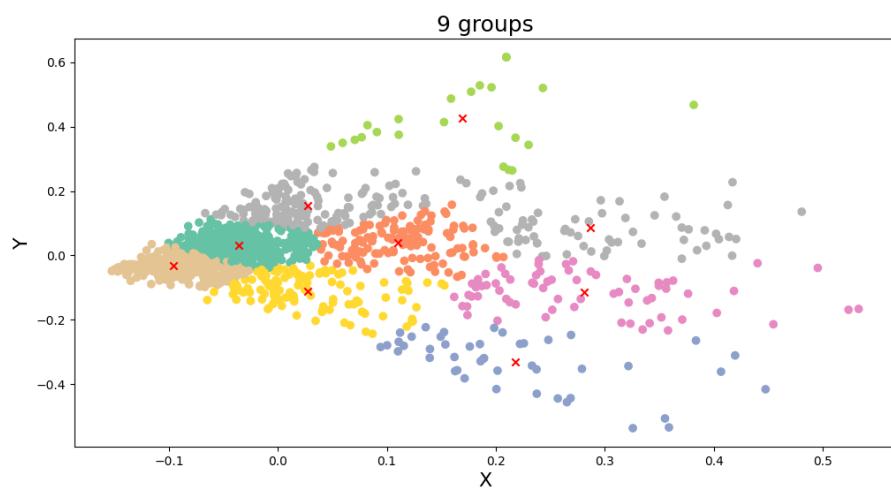
การจัดกลุ่มของข้อมูลหรือการทำ Text clustering นั้นโครงงานนี้จะใช้เทคนิค K-Means มาใช้ในการจัดกลุ่มโดยเฉลี่ยจำนวนกันกลุ่มจากการทำ Elbow method เพื่อหาจำนวนกลุ่มที่ดีที่สุดได้ตั้งภาพที่ 32 โดยจำนวนกลุ่มที่เหมาะสมจะอยู่ที่บริเวณส่วนโคงคล้ายข้อศอกในที่นี้จะประมาณกลุ่มได้ 6-9 กลุ่ม



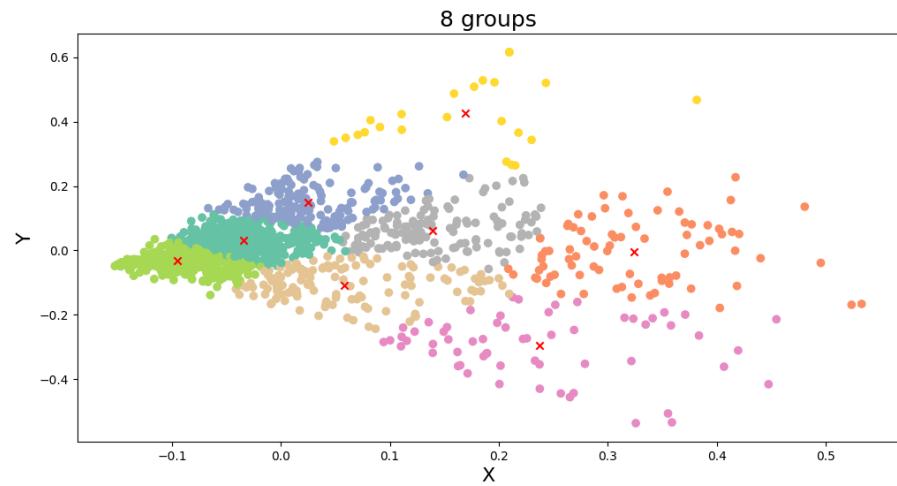
ภาพที่ 32 การทำ Elbow method

จากนั้นทำการทดลองจัดกลุ่มข้อมูลด้วยจำนวนกลุ่มที่แตกต่างกันแต่ข้อมูลเดียวกัน ทำการทดลองของจัดกลุ่มจำนวน 4 กลุ่มได้ดังนี้

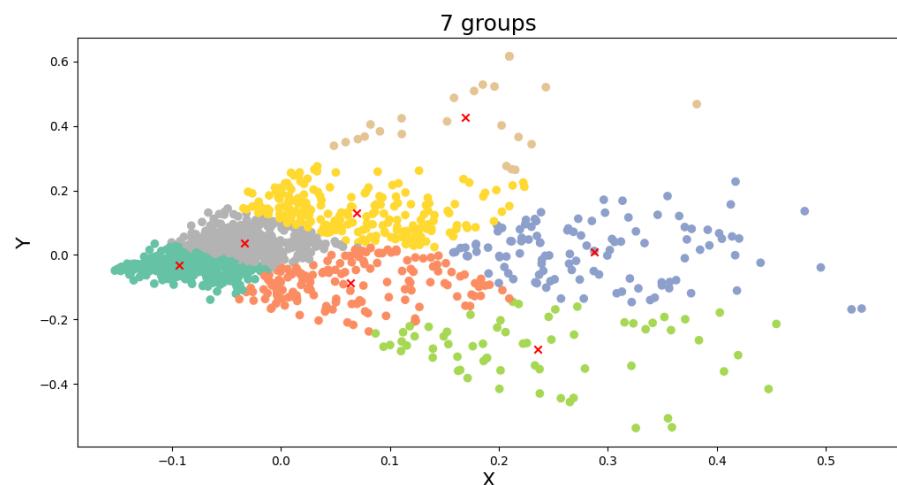
1. จำนวน 9 กลุ่ม
2. จำนวน 8 กลุ่ม
3. จำนวน 7 กลุ่ม
4. จำนวน 6 กลุ่ม



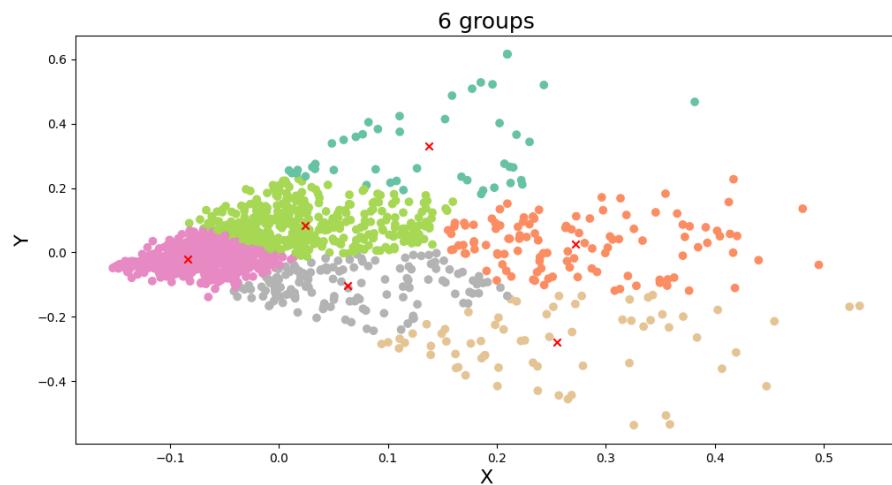
ภาพที่ 33 จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม



ภาพที่ 34 จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม



ภาพที่ 35 จัดกลุ่มข้อมูลจำนวน 7 กลุ่ม



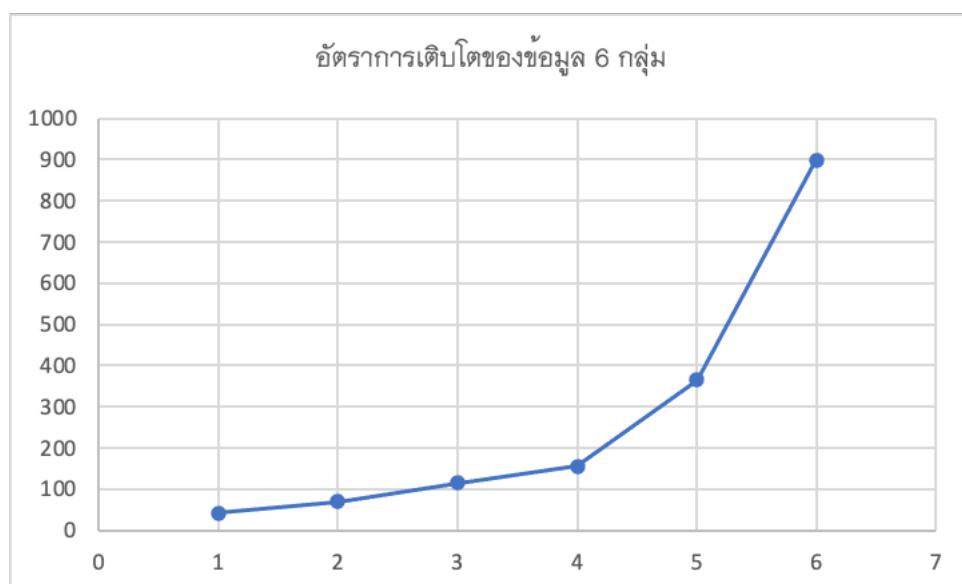
ภาพที่ 36 จัดกลุ่มข้อมูลจำนวน 6 กลุ่ม

ทำการทดลองนับจำนวนรายการบริษัทแต่ละประเภทในแต่ละกลุ่มเพื่อค่าเฉลี่ยและการกระจายตัวของข้อมูลบริษัท

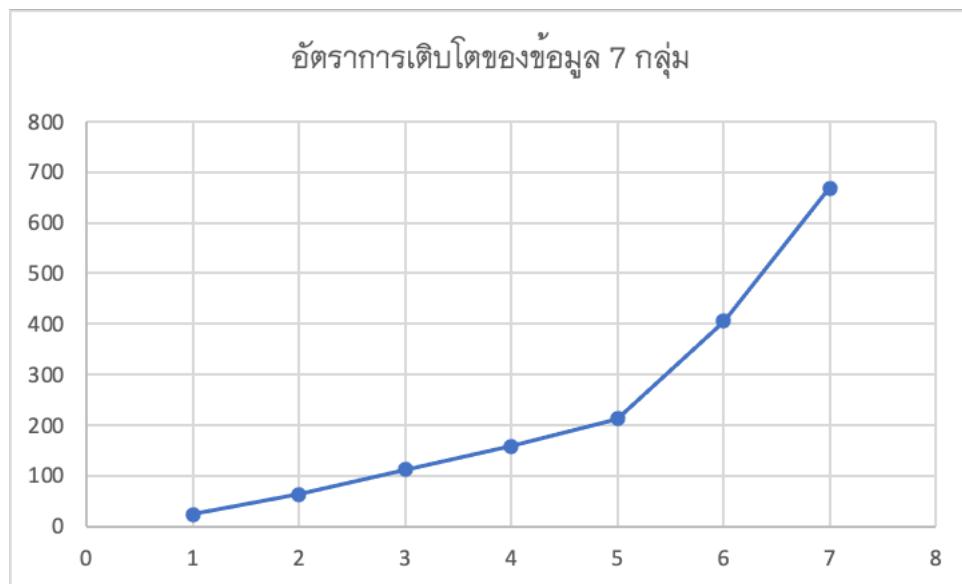
ตาราง 15 แสดงการนับจำนวนบริษัทแต่ละประเภทในการจัดกลุ่ม

จำนวนกลุ่ม	ประเภทของบริษัท	ค่าเฉลี่ย
6	<ul style="list-style-type: none"> - กลุ่ม 1 มี 115 รายการ - กลุ่ม 2 มี 899 รายการ - กลุ่ม 3 มี 69 รายการ - กลุ่ม 4 มี 155 รายการ - กลุ่ม 5 มี 41 รายการ - กลุ่ม 6 มี 364 รายการ 	469.43
7	<ul style="list-style-type: none"> - กลุ่ม 1 มี 405 รายการ - กลุ่ม 2 มี 24 รายการ - กลุ่ม 3 มี 112 รายการ - กลุ่ม 4 มี 158 รายการ - กลุ่ม 5 มี 669 รายการ - กลุ่ม 6 มี 63 รายการ - กลุ่ม 7 มี 212 รายการ 	234.71
8	<ul style="list-style-type: none"> - กลุ่ม 1 มี 403 รายการ - กลุ่ม 2 มี 81 รายการ - กลุ่ม 3 มี 655 รายการ - กลุ่ม 4 มี 155 รายการ - กลุ่ม 5 มี 138 รายการ - กลุ่ม 6 มี 24 รายการ - กลุ่ม 7 มี 131 รายการ - กลุ่ม 8 มี 56 รายการ 	205.38

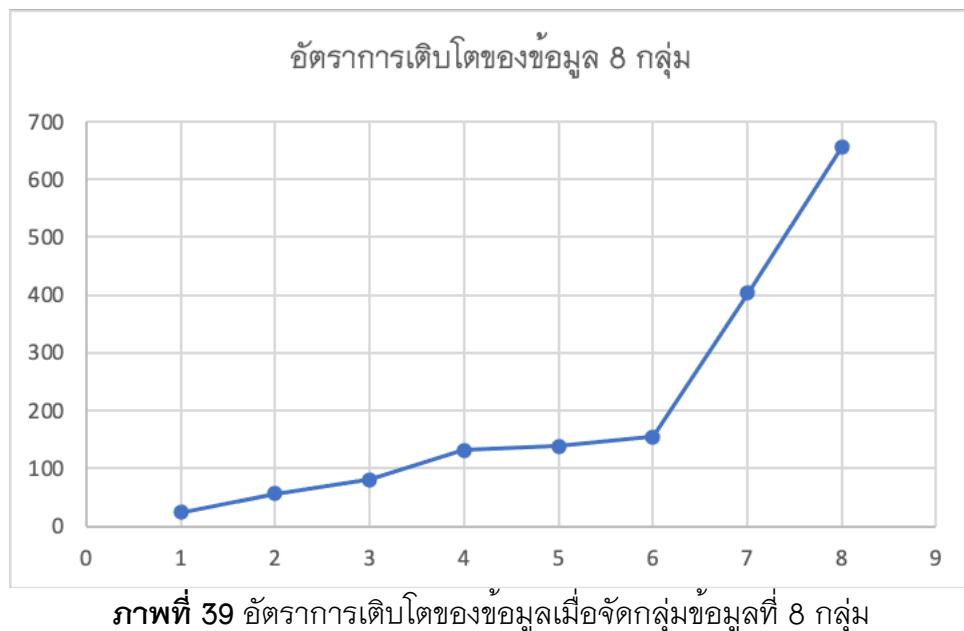
9	<ul style="list-style-type: none"> - กลุ่ม 1 มี 391 รายการ - กลุ่ม 2 มี 69 รายการ - กลุ่ม 3 มี 640 รายการ - กลุ่ม 4 มี 68 รายการ - กลุ่ม 5 มี 45 รายการ - กลุ่ม 6 มี 121 รายการ - กลุ่ม 7 มี 139 รายการ - กลุ่ม 8 มี 146 รายการ - กลุ่ม 9 มี 24 รายการ 	182.56
---	---	--------



ภาพที่ 37 อัตราการเติบโตของข้อมูลเมื่อจัดกลุ่มข้อมูลที่ 6 กลุ่ม



ภาพที่ 38 อัตราการเติบโตของข้อมูลเมื่อจัดกลุ่มข้อมูลที่ 7 กลุ่ม



จากภาพที่ 33–36 จะเห็นได้ว่าการกระจายตัวของข้อมูลนั้นมีความแตกต่างกันสังเกตได้จากสีที่ระบุตำแหน่งของกลุ่มในแต่ละ卦 โดยเฉพาะเมื่อเปรียบเทียบการแบ่งกลุ่มที่ 6 และ 8 จะเห็นว่าขอบเขตของข้อมูลของการแบ่ง 6 กลุ่มนั้นมีความกว้างมากและห่างไกลจากจุดกึ่งกลางของข้อมูล (Centroids point) ซึ่งประมาณได้ว่าการแบ่งกลุ่มที่จำนวน 6 กลุ่มนั้นอาจไม่ได้ประสิทธิภาพความแม่นยำมากพอก และในตรงกันข้ามการแบ่งกลุ่มข้อมูลที่ 9 กลุ่มนั้นจะเห็นได้ว่าขอบเขตของข้อมูลนั้นเล็กมากจนถึงทับซ้อนกันในแต่ละกลุ่มถึงแม้ขอบเขตของข้อมูลจะอยู่ใกล้จุดกึ่งกลางของข้อมูล ดังนั้นผู้วิจัยจึงตัดตัวเลือกการแบ่งกลุ่มข้อมูลที่ 6 กลุ่มออกเหลือเพียง 7 8 และ 9 กลุ่ม

และจากภาพ 37-40 จะเห็นได้ว่าการเติบโตของข้อมูล เมื่อนำมาปริมาณบริษัทในแต่ละประเภท มาเรียงจากน้อยไปมากนั้น อัตราการเติบโตของ การจัดกลุ่มข้อมูลที่ 7 กลุ่มนั้นมีอัตราการเติบโตที่คงที่มากที่สุด

จากนั้นทดลองทำการสุ่มเรียกข้อมูลในแต่ละกรณีของมาเพื่อประกอบการตัดสินใจว่าจะเลือกจำนวนกลุ่มของข้อมูลดังภาพด้วยค่าที่ 37-39

```
onze@Tinngritis-MacBook-Pro:~/desktop/_final_project
```

```
python main.py develop |1 ?2
```

```
detail: บริการรับออกแบบและพัฒนาระบบเว็บไซต์เพื่อธุรกิจต่างๆ (Web & Mobile Development)
cluster: 1

detail: บริการรับออกแบบและพัฒนาฐานข้อมูลเบื้องต้น สำหรับ บริษัท ราชการ หน่วยงาน ฯ ตามความต้องการไฟฟ้าทุกรูปแบบ
cluster: 2

detail: บริการพัฒนาซอฟต์แวร์เชิงโปรแกรมเพื่อการอ่านเรียนภาษา ภาษาต่างๆ ให้ได้ปรับเปลี่ยนตามภาษาต่างๆ ภาษาไทย อังกฤษ ติดต่อ ประมวลผลภาษา ภาษาอังกฤษ อ่านภาษาอังกฤษ และภาษาอื่นๆ เช่นภาษาจีนภาษาญี่ปุ่นภาษาฝรั่งเศส ฯลฯ
cluster: 3

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย ภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 4

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย ภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 5

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย ภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 6

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย ภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 7
```

ภาพที่ 41 ตัวอย่างการสุ่มข้อมูลของ การจัดกลุ่มข้อมูลที่ 9 กลุ่ม

```
onze@Tinngritis-MacBook-Pro:~/desktop/_final_project
```

```
python main.py develop |1 ?2
```

```
detail: บริการเรียนรู้ภาษาต่างๆ และพัฒนา Web Application เพื่อให้ได้รับการจัดการข้อมูลต่างๆ บนเว็บไซต์ได้ถูกต้อง เช่น ใช้งานภาษาต่างประเทศ W3C เพื่อการเผยแพร่ต่อสื่อสาร Web Browser โซเชียลมีเดีย ฯลฯ จัดทำข้อมูลและอัปเดตให้กับผู้ใช้งาน
cluster: 1

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย และภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 2

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย และภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 3

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย และภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 4

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย และภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 5

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย และภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 6

detail: บริการสอนภาษาต่างๆ เช่นภาษาไทย และภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น ภาษาฝรั่งเศส ฯลฯ รวมถึงภาษาอิตาลี ภาษาเยอรมัน ภาษาสเปน ภาษาฝรั่งเศส ฯลฯ
cluster: 7
```

ภาพที่ 42 ตัวอย่างการสุ่มข้อมูลของ การจัดกลุ่มข้อมูลที่ 8 กลุ่ม

```
onze@Tinngritis-MacBook-Pro:~/desktop/_final_project
```

```
python main.py develop |1 ?2
```

```
detail: บริการรับเขียนภาษา จัดทำ ออกแบบ ระบบ Construction Management และ Facilities Management, IoT
cluster: 0

detail: บริการด้าน Digital Marketing Agency (Content Marketing, Photography & Video, Signage, Collateral, Digital, Branding)
cluster: 1

detail: บริการ Platform ชุมชนธุรกิจสู่ธุรกิจ SMEs ที่เน้นไปทางการขายอิเล็กทรอนิกส์ On Cloud และ โปรดักส์อิเล็กทรอนิกส์
cluster: 2

detail: บริการระบบขายอิเล็กทรอนิกส์ (Digital Marketplace) จัดซื้อขาย อิเล็กทรอนิกส์ (Credit Access) จัดทำ洞察การวิเคราะห์ (Insights & Analytics) แพลตฟอร์มขายและปริมาณข้อมูล
cluster: 3
```

ภาพที่ 43 ตัวอย่างการสุ่มข้อมูลของ การจัดกลุ่มข้อมูลที่ 7 กลุ่ม

จากการทดลองสุ่มเรียกข้อมูลหลาย ๆ ครั้งพบว่าจำนวนของการจัดกลุ่มที่มีค่าข้อมูลทับช้อนกันน้อยที่สุดอยู่ที่ 7 กลุ่มทำให้ผู้วิจัยเลือกที่จะแบ่งกลุ่มข้อมูลที่ 7 กลุ่มแต่ทั้งนี้ก็ยังมีข้อมูลที่ทับช้อนกันอยู่บ้างเล็กน้อยซึ่งอยู่ในระดับที่รับได้ และเมื่อทำการแบ่งกลุ่มข้อมูลเรียบร้อยแล้วก็สามารถบันทึกข้อมูลพร้อมกับ ID ของกลุ่มเพื่อนำไปนำเข้าฐานข้อมูลและใช้งานต่อไป

4.1.3 การกำหนดชื่อกลุ่ม

เมื่อได้ข้อมูลที่สมบูรณ์อยู่ในฐานข้อมูลแล้วนั้นการแสดงผลข้อมูลของกลุ่มจากหน้าเว็บไซต์จำเป็นต้องมีการตั้งชื่อกลุ่มนึ่งจากข้อมูลที่ได้จากการจัดกลุ่มคือ ID ซึ่งคือตัวเลขตั้งแต่ 0-6 เนื่องจากทำการกำหนดจำนวนกลุ่มไว้ที่ 7 กลุ่มดังนั้นเพื่อให้การแสดงผลในหน้าเว็บไซต์และให้การเรียกกลุ่มงานขึ้นจึงทำการตั้งชื่อกลุ่มโดยชื่อจะอยู่ในประเภทของเทคโนโลยีเนื่องจากข้อมูลที่มีนั้นเป็นบริษัทที่ทำเกี่ยวกับอุตสาหกรรมเทคโนโลยีหรือบริษัททางด้านไอที การตั้งชื่อของกลุ่มข้อมูลนั้นได้ทำการข้างขึ้นมาจากประเทงานโดยที่จากเว็บไซต์ th.jobsdb.com เป็นหลักซึ่งมีอยู่ 18 ประเภท รายชื่อประเทงานโดยที่มีในเว็บไซต์ th.jobsdb.com มีดังนี้

1. งาน Application Network
2. งาน Software
3. งาน Database
4. นักวิทยาศาสตร์ข้อมูล
5. งาน Hardware
6. งาน IT Audit
7. งานปรีกษาโฉม
8. งาน IT Project
9. งานดูแลเว็บไซต์ งาน SEO
10. งาน MIS
11. งาน Mobile งาน Wireless communications
12. งานดูแลระบบ Network
13. งานโปรแกรมเมอร์
14. งาน IT Security
15. งาน IT Support
16. งาน Software Tester
17. นักออกแบบ UI/UX
18. งานโอดีอีน ๆ

แต่ในโครงงานนี้มีกลุ่มข้อมูลเพียง 7 กลุ่มดังนั้นจึงต้องเลือกประเทงานที่ตรงกับข้อมูลในกลุ่มมากที่สุดเท่านั้น โดยวิธีที่ใช้เลือกคือการสุ่มข้อมูลในแต่ละกลุ่มตั้งแต่ 0-6 มาและตรวจดูว่าควรจะได้ชื่อกลุ่มเป็นประเทงานไหนดังภาพด้าน右ที่ 40-41

```

onze@Tinngrits-MacBook-Pro:~/desktop/final_project
.../final_project (-zsh) 6.6 GB .../final_project (-zsh) .../Desktop/final_project 20:24:16
python cluster_naming.py
Group: 0
1: บริการ Platform ซอฟต์แวร์ที่บันทึกการดำเนินการทางธุรกิจในร้านค้าและขายของออนไลน์
2: บริการลักษณะเดียวกับ Hospitality Industry Management Systems (Front Office, Point of Sale, Call Accounting, Back Office with Accounts Payable, Accounts Receivable, General Ledger and Inventory Control)
3: บริการแพลตฟอร์มการจัดซื้อซึ่งรวมเครื่องมือและเอกสารในการจัดซื้อและจัดจ้าง
4: บริการแบบ Paperless Solution และการจัดการเอกสาร Dform Paperless System ที่พัฒนา Cloud based (คือค้นหาน้ำหนังในระบบ) และ On-Premise (คือหันวิเคราะห์ข้อมูล)
5: บริการ Software as a Service on Cloud รวมทั้งการจัดการห้องเรียนและห้องเรียน
6: บริการแพลตฟอร์ม VIRTUAL REALITY (VR) VR Sphere รวมพื้นที่ด้วยเทคโนโลยี VR, Software, Hardware, Mechanic หรือ Product Design, รวม Smart Workplace Solution (Meet in Touch, CO Desk, Visitor) (Visitor Management System, Hot Desk System, Meeting Room Reservation System, Locker, On-Premise Face Recognition System)
7: บริการ Software as a Service ที่เกี่ยวกับการเฝ้าระวังผู้ใช้งานเช่น Big Data หรือบันทึกผู้ใช้งานเพื่อการตัดสินใจ (Social Monitoring, Social Engagement, Web-based social intelligence solution, integrating social listening, monitoring, engagement and customer service)
8: บริการระบบอิเล็กทรอนิกส์สำหรับห้องเรียนและห้องเรียน (E-Library) และนี้มีตัวอย่างเช่น Auto-Lib และนี้มีตัวอย่างเช่น และระบบ Cloud รวมถึง Digital Knowledge Center, Learning Management System
9: บริการแพลตฟอร์มที่ให้บริการห้องเรียนและห้องเรียน เช่น Digital Infrastructure หรืออื่นๆ เช่น Ascend Honey, Fintech & E-Payments: TrueMoney, Ascend Nano
Ascend Commerce: E-Commerce / E-Procurement / Fulfillment: WeLoveShopping, WeMall, WeFresh, GoodChoiz.com, Pantavani, Ascend Travel, Egg Digital, Aden Digital Enablers: Data Center & Cloud Service Provider: TrueIDC
10: บริการซอฟต์แวร์ที่ควบคุมชิปเซ็ต ไมโครคอนโทรลเลอร์ (Micro-Controller) FiO และ RapidSTM32 Blockset หรือแม่บอร์ดพิเศษ (Custom Blockset Development), Smart Sensor

```

ภาพที่ 44 ตัวอย่างข้อมูลในกลุ่ม 0

```

onze@Tinngrits-MacBook-Pro:~/desktop/final_project
.../final_project (-zsh) 6.6 GB .../final_project (-zsh) .../Desktop/final_project 20:25:32
python cluster_naming.py
Group: 2
1: บริการด้านการตลาด ออนไลน์ บริการด้านเว็บไซต์, Digital Marketing, Web Design, Analytics & Web Analysis, Online Advertising, Social Media Marketing, SEO/SME/ASO, Mobile App Development, Mobile Marketing, Cloud Solution, E-books Development, iBeacons, Digital Marketing Workshops
2: บริการแพลตฟอร์มที่รวม Software Development, Graphic Design and Web Design, WEB DESIGN & WEB PROGRAMMING, SEO search engine optimizer
3: บริการด้านการดีไซน์ UX/UI design, API connections, e-commerce system & payment gateway, Digital Tracking & Analytics
4: บริการด้านการแพทย์ ห้องปฏิบัติการ ห้องผ่าตัด ห้องผู้ป่วย ห้องแม่บ้าน ห้องน้ำ ห้องน้ำส่วนตัว ห้องน้ำส่วนตัว (Digital Solution Consulting Service, Medical Software, Develop Web / Mobile Application)
5: บริการ SEO, Web Design, Digital Marketing
6: บริการด้าน digital marketing agency
7: บริการด้านเว็บไซต์ที่มีความเฉพาะเจาะจง เช่น Web Application, Web Hosting E-Commerce, SEO (Search Engine Optimisation) and website maintenance., เพื่อปรับปรุงให้กับสิ่งที่ต้องการ online marketing
8: บริการด้านเว็บไซต์ที่มีความเฉพาะเจาะจง เช่น Web Design & Development, Website Promotion, Webbased Application, Website Branding, Domain Name Registration, Hosting Solution for E-Business Applications, E-Commerce Solution, Search Engine Consultation, User Interface Design, Usability Testing, Shopping Cart, Payment Gateway
10: บริการด้านเว็บไซต์ที่มีความเฉพาะเจาะจง เช่น Web Application, Software Development, Platform Application) a startup software development studio Web App, Application, Design

```

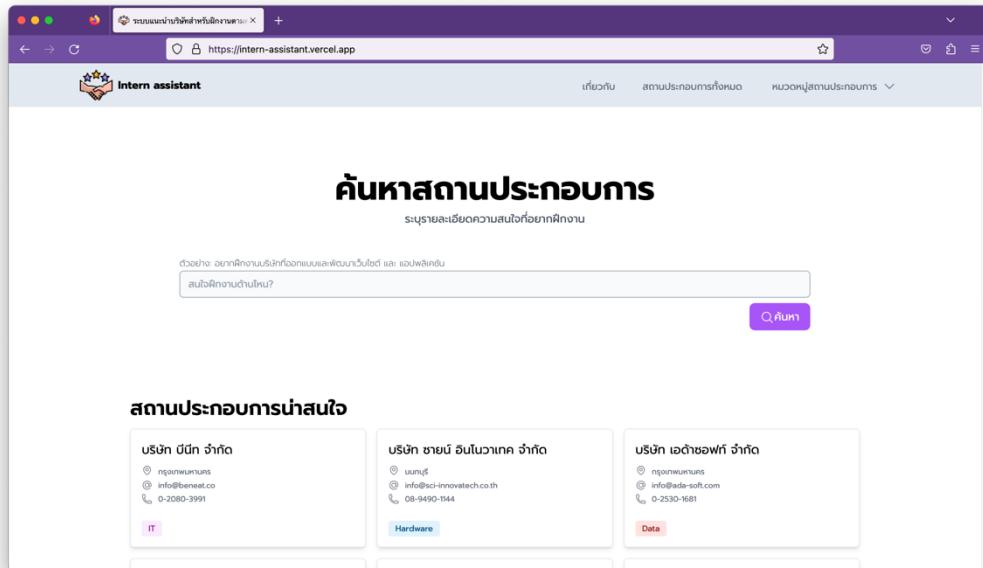
ภาพที่ 45 ตัวอย่างข้อมูลในกลุ่ม 2

จากการทำการสุ่มเรียกข้อมูลดูทั้ง 7 กลุ่ม รายชื่อประเภทงานที่สามารถใช้ตั้งชื่อกลุ่มข้อมูลโดยอ้างอิงของเว็บไซต์ th.jobsdb.com มีดังนี้

1. Data analysis
2. Online marketing
3. Software
4. Hardware
5. Network
6. IT
7. Other

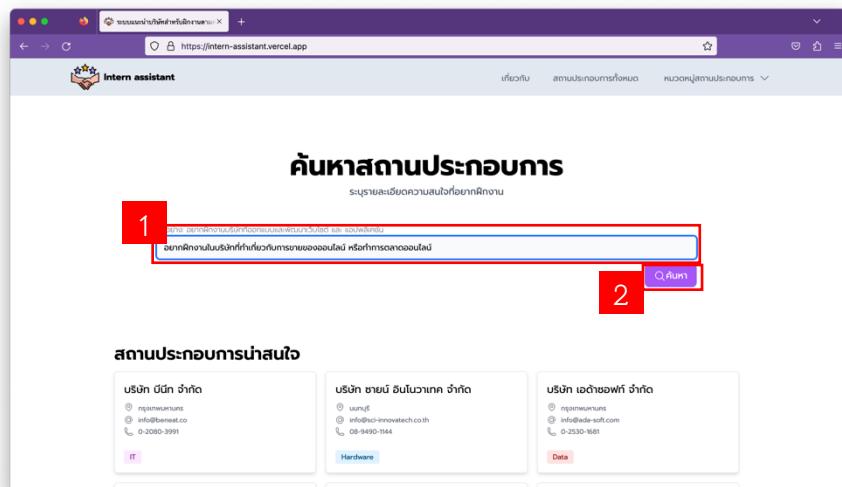
4.2 ขั้นตอนการใช้งานสำหรับผู้ใช้งาน

4.2.1 หน้าแรกเว็บไซต์ Intern–assistant



ภาพที่ 46 หน้าแรกเว็บไซต์ Intern–assistant

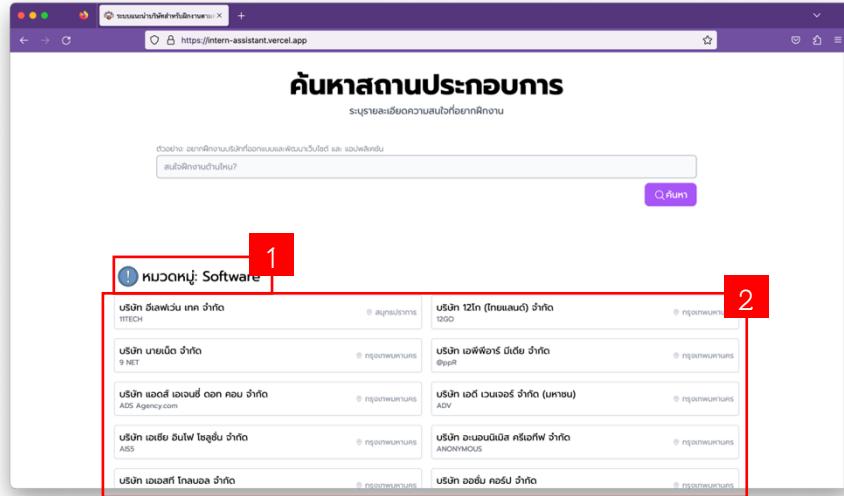
4.2.2 ค้นหาบริษัท



ภาพที่ 47 ค้นหาบริษัท

หมายเลขอ 1 พิมพ์ค้นหาบริษัทที่ต้องการฝึกงานด้วยรายละเอียดความสนใจรูปแบบธุรกิจ
หมายเลขอ 2 กดปุ่มค้นหา

4.2.3 หน้าแสดงผลลัพธ์การค้นหา^ช

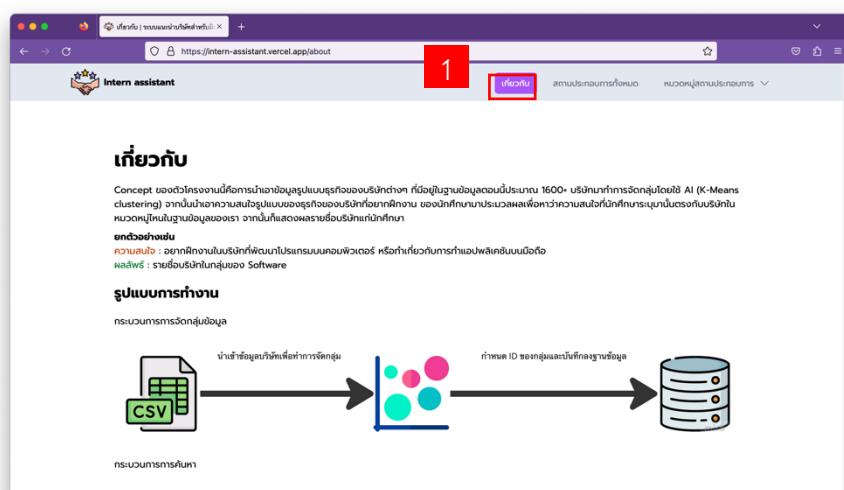


ภาพที่ 48 หน้าแสดงผลลัพธ์การค้นหา^ช

หมายเหตุ 1 ชื่อของกลุ่มข้อมูล

หมายเหตุ 2 รายชื่อบริษัททั้งหมดที่มีอยู่ในกลุ่ม

4.2.4 หน้าเกี่ยวกับ



ภาพที่ 49 หน้าเกี่ยวกับ

หมายเหตุ 1 เมนูหน้าเกี่ยวกับ

4.2.5 หน้าแสดงรายชื่อบริษัททั้งหมด

1

2

บริษัท สุรินทร์วัน จำกัด	บริษัท 10 กันยา จำกัด	บริษัท วีโนเว่น ไทยเน็ต จำกัด
บริษัท 11-инфอร์มิก ดิจิทัล จำกัด	บริษัท ลีฟท์แอนด์ เทค จำกัด	บริษัท 12Go (ไทยแลนด์) จำกัด
บริษัท วันไกรสิน นิคมอุตสาหกรรม จำกัด	สำนักงานสถาบันฯ จำกัด 147 KP Studio Ltd., Part	บริษัท วันดี จำกัด One DD Hub Co., Ltd.
บริษัท วัน มีเดียซอฟต์ เทคโนโลยี จำกัด	บริษัท วันโซลูชัน โซลูชันดิจิทัล จำกัด 1 Hotel Solutions.com Co., Ltd.	บริษัท วีซีพี (ประเทศไทย) จำกัด 2C2P (Thailand) Co., Ltd.
บริษัท กฎ ใจดี แอนด์ เอฟเฟกต์ จำกัด	บริษัท 360 บันดิวนฟ์ จำกัด 360 Innovative Co., Ltd.	บริษัท นีวูวู จำกัด 3View Group Co., Ltd.
บริษัท ก้าว วันเพล็ก (ประเทศไทย) จำกัด	บริษัท ไฟร์ ศิลป์แอร์ จำกัด	บริษัท พลังเมือง จำกัด

ภาพที่ 50 หน้าแสดงรายชื่อบริษัททั้งหมด

หมายเหตุ 1 เมนูหน้ารายชื่อบริษัททั้งหมด

หมายเหตุ 2 รายชื่อบริษัททั้งหมด

4.2.6 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

1

2

3

หมายเหตุ: Data

บริษัท 11-инфอร์มิก ดิจิทัล จำกัด	บริษัท 9 Net จำกัด
บริษัท ก้าว วันเพล็ก (ประเทศไทย) จำกัด	บริษัท นายดิบ จำกัด
บริษัท ดีโซลาร์ คอมพิวเตอร์ (ประเทศไทย) จำกัด	บริษัท เอ-ไฮส์ จำกัด
บริษัท แมท เอเช ทัชเน็ต แอนด์ แมกซ์พีเพลทเวอร์ก จำกัด	บริษัท และวันเนอร์ บีชิพ ไอซ์บีส์ คอมปะร์เน็น จำกัด
บริษัท แมตเต็กเกอร์ คอมปะร์เน็น จำกัด	บริษัท อดาสตรา (ไทยแลนด์) จำกัด
บริษัท อาดี จำกัด	บริษัท อะดีเพรส จำกัด
บริษัท อเมริกา คอมพิวเตอร์ เอเชีย จำกัด	บริษัท แอร์พอร์ตส์ จำกัด

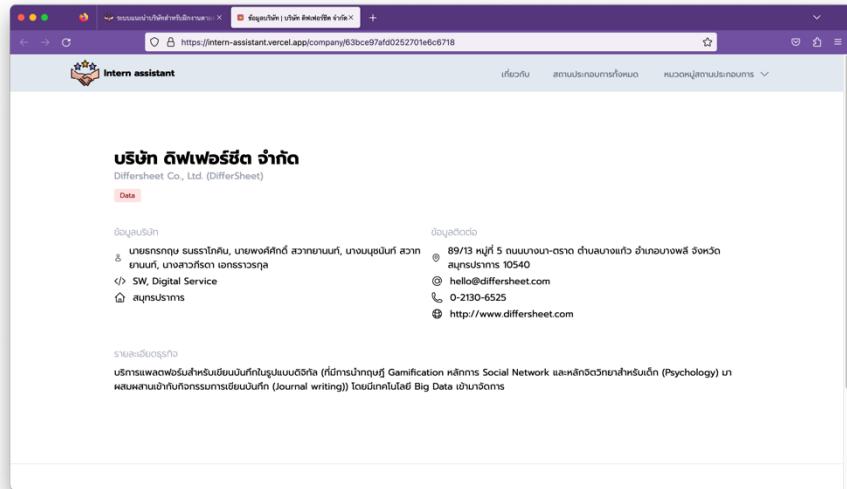
ภาพที่ 51 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หมายเหตุ 1 เมนูรายชื่อกลุ่มบริษัท

หมายเหตุ 2 รายชื่อกลุ่มบริษัท

หมายเหตุ 3 รายชื่อบริษัทในกลุ่ม

4.2.7 หน้ารายละเอียดบริษัท



ภาพที่ 52 หน้ารายละเอียดบริษัท

4.3 การวัดค่าความคล้ายคลึง

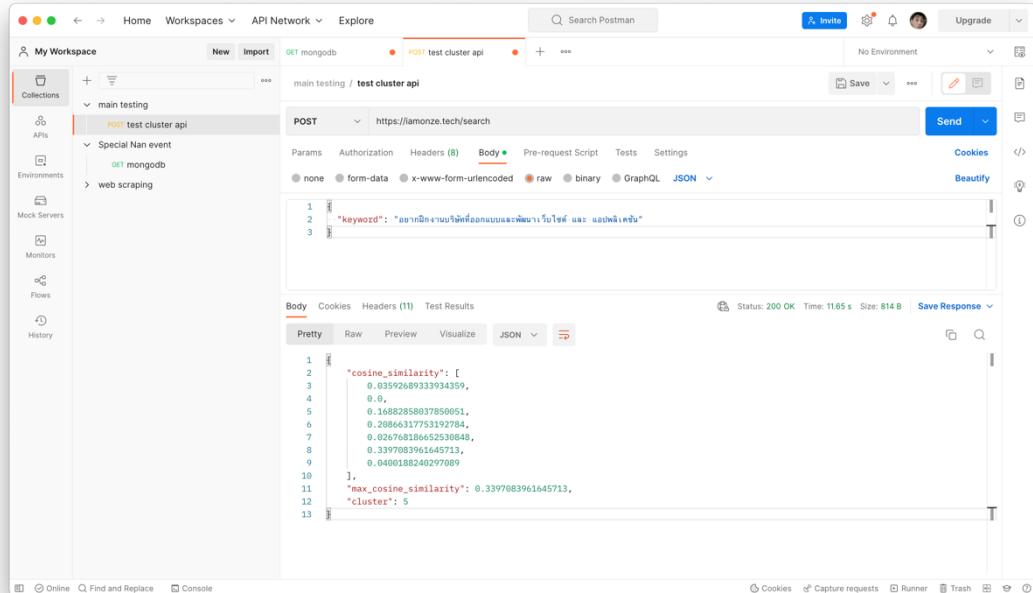
Cosine similarity เป็นเทคนิคที่นำมาใช้หาความคล้ายคลึงระหว่างความสนใจของผู้ใช้และข้อมูลบริษัทที่อยู่ในฐานข้อมูลยิ่งค่า Cosine similarity เข้าใกล้ 1 แสดงว่าประโยชน์นี้มีความคล้ายคลึงกับข้อมูลบริษัทในกลุ่มนั้นมากดังภาพที่ 49

```
onze@Tinngrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh) ● 1 ..final_project (-zsh) ⇧ 2 + 
20% 4.4 GB develop + ~/Desktop/final_project
└ python cosine_similarity.py
Keyword: ออกแนวโน้มที่ตัวเอง react js ทำให้เกิดการเขียนเว็บ การตลาดออนไลน์ตัวเอง และ SEO
cluster: 6
cosine similarity: 0.07871941760441518
~/desktop/final_project ➜ develop !3 ?2 7s 12:56:32
```

ภาพที่ 53 ตัวอย่างการคำนวณค่า Cosine similarity

4.3.1 การหากลุ่มที่มีความคล้ายคลึงมากที่สุด

- คำนวณค่า Cosine similarity ทุกกลุ่มจากความสนใจของผู้ใช้ผ่าน API ที่สร้างไว้เพื่อคำนวณโดยเฉพาะ



ภาพที่ 54 ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API

บทที่ 5

สรุปผลการดำเนินงาน

การพัฒนาระบบแนะนำบริษัทสำหรับผู้ใช้งานตามความสนใจ ด้วยเทคโนโลยีปัญญาประดิษฐ์ ในครั้งนี้สามารถสรุปการดำเนินงาน ปัญหาที่เกิดขึ้นระหว่างการดำเนินงาน ขอเสนอและแนวทางพัฒนาต่อไปดังนี้

5.1 สรุปผลการดำเนินงาน

จากการการดำเนินงานได้นำข้อมูลบริษัทจากสมาคมปัญญาประดิษฐ์แห่งประเทศไทยมาทำการจัดกลุ่มโดยใช้เทคนิคการจัดกลุ่มข้อมูล (K-Means) และใช้เทคนิคการคำนวณค่าความคล้ายคลึง (Cosine similarity) ในการหาความคล้ายคลึงของความสนใจรูปแบบงานของผู้ใช้กับข้อมูลบริษัทที่ได้ผลสรุปดังนี้

การจัดกลุ่มข้อมูลทั้งหมดจำนวน 4 กลุ่มได้แก่ การจัดกลุ่มข้อมูลที่ 6 7 8 และ 9 กลุ่มได้ผลสรุปดังนี้ กรณีแบ่งกลุ่มข้อมูลที่ 6 กลุ่มนี้มีอสูรเมืองเรียกข้อมูลดูแล้วพบว่าขอบเขตของกลุ่มนี้กว้าง เกินไปและมีข้อมูลทับซ้อนกันจำนวนมาก กรณีแบ่งข้อมูลที่ 7 กลุ่มพบว่าเมื่อสูรเมืองเรียกดูข้อมูล ข้อมูลมีความทับซ้อนกันน้อยมากและขอบเขตของข้อมูลก็อยู่ในระดับที่เหมาะสมสมยอมรับได้ กรณีแบ่งกลุ่มที่ 8 กลุ่มพบว่ามีความคล้ายเคียงกับการแบ่งกลุ่มที่ 7 กลุ่มแต่ขอบเขตของข้อมูลบางกลุ่มนั้นแคบ เกินไปทำให้มีเนื้อหาที่ซ้ำกันกับกลุ่มอื่น และกรณีแบ่งกลุ่มข้อมูลที่ 9 กลุ่มพบว่าขอบเขตของข้อมูลนั้นแคบที่สุดและแต่ละกลุ่มนั้นมีความทับซ้อนกันค่อนข้างมากจึงเกิดกลุ่มที่มีเนื้อหาแบบเดียวกันแต่อยู่คนละกลุ่ม

ซึ่งจะเห็นได้ว่ากรณีการแบ่งกลุ่มที่ 6 และ 9 กลุ่มนี้ขอบเขตของเนื้อหานั้นอยู่ในระดับที่ไม่ค่อยติดกันเมื่อเทียบกับการแบ่งกลุ่มที่ 7 และ 8 กลุ่ม และในการทดสอบองริช Elbow method จุดที่อยู่ตรงมุมอยู่ระหว่างจุดที่ 7-8 ผู้วิจัยจึงเลือกเปรียบเทียบกันและหาขอสรุปได้ว่าเลือกแบ่งกลุ่มที่ 7 กลุ่มเป็นการแบ่งกลุ่มข้อมูลที่มีประสิทธิภาพมากที่สุดจากการสูรเมืองเรียกดูข้อมูลของแต่ละกรณีเพื่อมาใช้ในการพัฒนาระบบ

ในการหาความคล้ายคลึงของข้อมูลด้วยวิธี Cosine similarity นั้นผลลัพธ์การคำนวณและการวิเคราะห์เพื่อหากลุ่มที่เหมาะสมกับความสนใจของผู้ใช้นั้นพบว่าเมื่อได้กลุ่มจากที่การคำนวณแล้วนั้นรูปแบบธุรกิจค่อนข้างตรงกับความสนใจที่ผู้ใช้ลงมา

5.1.1 จุดเด่นของระบบ

1. ขั้นตอนการใช้งานของผู้ใช้นักถูกออกแบบมาให้ใช้งานง่ายและสะดวกรวดเร็ว
2. ออกแบบหน้าจอแสดงผล (User interface) เข้าใจง่ายและใช้งานได้สะดวกไม่ซับซ้อน สามารถรองรับได้ทุกอุปกรณ์
3. ระบบสามารถเข้าถึงง่ายเนื่องจากพัฒนาอยู่ในรูปแบบของ Web application ทำให้ไม่ต้องติดตั้งก่อนใช้งานสามารถใช้งานผ่าน Browser ได้ในทุกอุปกรณ์

5.2 สรุปปัญหาที่เกิดระหว่างการดำเนินงาน

5.2.1 ในการที่จะเพิ่มข้อมูลบริษัทลงในฐานข้อมูลเพิ่มจำเป็นต้องทำการ Word segmentation ข้อมูลใหม่ที่จะเข้ามารวมกับข้อมูลเดิมที่มีอยู่แล้วต้องนั่งข้อมูลของแต่ละบริษัทที่จะถูกแปลงเป็นภาษาไทยไปทุกครั้งทั้งมีการเพิ่มข้อมูลใหม่

5.2.2 เมื่อทำการ Clustering ข้อมูลใหม่แล้วต้องทำการตั้งใช้ให้กับกลุ่มข้อมูลใหม่ เพราะเมื่อมีข้อมูลที่เปลี่ยนไปเนื่องจากลุ่มเดิมก็อาจเปลี่ยนไปยกตัวอย่างเช่น กลุ่มที่ 0 เดิมเป็นกลุ่มของ Network แต่เมื่อมีการเพิ่มข้อมูลใหม่และทำการ Clustering ใหม่กลุ่ม 0 ก็อาจจะกลายเป็น Data เพราะเนื้อหาในกลุ่มนั้นเปลี่ยนไป หรืออาจมีกลุ่มเพิ่มเติมขึ้นมาแทนหนึ่งจากปัจจุบัน

5.2.3 ข้อมูลที่ได้รับมาเมื่อมีคำที่สอดคล้องไม่สามารถแก้ไขได้ด้วยผู้พัฒนาเอง การทำ Clustering นั้นอาจจะไม่ได้แยกแยะข้อมูลได้ดียกตัวอย่างเช่น ประโยชน์ที่มีคำว่า “แอปพลิเคชัน” อาจจะอยู่คู่คลุ่มกับประโยชน์ที่มีคำว่า “แอพพลิเคชัน”

5.3 แนวทางพัฒนาระบบในอนาคต

- 5.3.1 พัฒนาความแม่นยำในการจัดกลุ่มข้อมูล
- 5.3.2 พัฒนาให้สามารถแนะนำตำแหน่งงานในบริษัทได้
- 5.3.3 เพิ่มชุดข้อมูลให้มากขึ้นเพื่อเพิ่มความแม่นยำของเว็บไซต์
- 5.3.4 พัฒนาความเร็วของอัลгорิทึมในการคำนวณความคล้ายคลึง (Cosine similarity)
- 5.3.5 พัฒนาขั้นตอนการเรียกดูข้อมูลให้ง่ายขึ้น
- 5.3.6 พัฒนาให้รองรับการกรองตัวเลือกที่จะคนหาก
- 5.3.7 พัฒนาเว็บไซต์ให้มีความปลอดภัยมากขึ้น

5.4 แบบประเมินความพึงพอใจของผู้ใช้

จากการประเมินความพึงพอใจของผู้ใช้ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ (Computer Internship Recommendation System With Artificial Intelligence Technology) สามารถสรุปผลได้ดังนี้

การกำหนดเกณฑ์การพิจารณา

เกณฑ์การพิจารณาระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ มีดังนี้

1. **เกณฑ์การให้คะแนน** ได้กำหนดเกณฑ์การให้คะแนนไว้ 5 ระดับดังนี้
 - 5 คะแนน หมายถึง ระดับความพึงพอใจ/ความเข้าใจดีมาก
 - 4 คะแนน หมายถึง ระดับความพึงพอใจ/ความเข้าใจดี
 - 3 คะแนน หมายถึง ระดับความพึงพอใจ/ความเข้าใจปานกลาง
 - 2 คะแนน หมายถึง ระดับความพึงพอใจ/ความเข้าใจน้อย
 - 1 คะแนน หมายถึง ระดับความพึงพอใจ/ความเข้าใจน้อยมาก
2. **เกณฑ์การแบ่งช่วงคะแนนค่าเฉลี่ย** เกณฑ์การแบ่งช่วงคะแนนค่าเฉลี่ยได้กำหนด
 - เกณฑ์การประเมินไว้ดังนี้
 - ค่าเฉลี่ย 4.51 – 5.00 หมายถึง ระดับความพึงพอใจ/ความเข้าใจดีมาก
 - ค่าเฉลี่ย 3.51 – 4.50 หมายถึง ระดับความพึงพอใจ/ความเข้าใจดี
 - ค่าเฉลี่ย 2.51 – 3.50 หมายถึง ระดับความพึงพอใจ/ความเข้าใจปานกลาง
 - ค่าเฉลี่ย 1.51 – 2.50 หมายถึง ระดับความพึงพอใจ/ความเข้าใจน้อย
 - ค่าเฉลี่ย 1.00 – 1.50 หมายถึง ระดับความพึงพอใจ/ความเข้าใจน้อยมาก

ตาราง 16 แสดงจำนวน ค่าเฉลี่ยความพึงพอใจต่อระบบ

รายละเอียด	ระดับความพึงพอใจ/ความเข้าใจ						เกณฑ์การประเมิน
	5	4	3	2	1	ค่าเฉลี่ย	
มีการออกแบบหน้าจอ สำหรับผู้ใช้งานอย่าง เหมาะสม	7	6	3	1	0	4.12	ดี
วิธีการใช้งานง่ายต่อการ ทำความเข้าใจ	6	7	3	1	0	4.06	ดี
ประสิทธิภาพความเสถียร ในการทำงานของเว็บแอป พลิเคชัน	2	9	6	0	0	3.76	ดี
ความเหมาะสมของ รูปแบบของหน้าจอแจ้ง เตือนต่าง ๆ	5	7	5	0	0	4	ดี
ความพึงพอใจในภาพรวม	5	9	3	0	0	4.12	ดี
ค่าเฉลี่ยรวม						4.01	ดี

จากการพบร่วมกับ “มีความพึงพอใจต่อระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ โดยภาพรวมอยู่ในเกณฑ์ประเมินที่ดี มีค่าเฉลี่ยเท่ากับ 4.01

ข้อเสนอแนะ

สามารถสรุปความพึงพอใจและข้อเสนอแนะของผู้ใช้ที่มีต่อ ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีปัญญาประดิษฐ์ ได้ดังนี้

ตาราง 17 ข้อเสนอแนะของผู้ใช้

ข้อที่	ข้อเสนอแนะ
1	ควรจะมีฟังก์ชันที่ให้เลือกจังหวัดว่าเรารออย่างใดของจังหวัดไหน เพื่อบางคนอยากทำงานใกล้บ้าน
2	อย่างให้มีการเลือกโซนของภาคอย่างภาคเหนือหรือภาคอีสานแต่โดยรวมทำได้ดีแล้ว
3	ต้องทำให้ผู้ใช้งาน ใช้งานได้ง่ายกว่านี้ และการค้นหาบ้างอย่างก็ไม่ตรงกับความต้องการที่คนหาเท่าไหร่

เอกสารอ้างอิง

- ตาเยะ, ช. (2022, 6 28). NLP คืออะไร . Retrieved from mindphp: <https://www.mindphp.com/คุณี/73-%E0%B8%95%E0%B8%97%E0%B8%9A%E0%B8%A1/8859-nlp.html>
- D'Agostino, A. (2021, 11 24). Text Clustering with TF-IDF in Python. Retrieved from medium: <https://medium.com/mlearning-ai/text-clustering-with-tf-idf-in-python-c94cd26a31e7>
- API คืออะไร. (n.d.). Retrieved from amazon: <https://aws.amazon.com/th/what-is/api/>
- บทความ E-R Diagram คืออะไร. (2014). Retrieved from 9experttraining: <https://www.9experttraining.com/articles/บทความ-e-r-diagram-%E0%B8%95%E0%B8%97%E0%B8%9A%E0%B8%A1>
- (2022, 3 12). Retrieved from Use Case Diagram: <https://www.mindphp.com/บทความ/31-%E0%B8%95%E0%B8%97%E0%B8%9A%E0%B8%A1/ทั่วไป/6870-use-case-diagram.html>
- การวิเคราะห์ระบบและการออกแบบ System Analysis and Design (ซิลสเต็ม อนาคต แอนด์ ดีไซน์). (2022, 3 30). Retrieved from mindphp: <https://www.mindphp.com/บทความ/31-%E0%B8%95%E0%B8%97%E0%B8%9A%E0%B8%A1/ทั่วไป/4084-system-analysis-and-design.html>
- mindphp. (n.d.). NumPy คืออะไร. Retrieved from mindphp: <https://www.mindphp.com/บทเรียนออนไลน์/83-python/8492-what-is-the-numpy.html>
- ทำความรู้จักกับ Node.js. (2021, 3 1). Retrieved from marcuscode: <http://marcuscode.com/tutorials/nodejs/introducing-nodejs>
- CSS คืออะไร ? มีประโยชน์อย่างไรบ้าง. (2016, 3 16). Retrieved from wynnsoft: <https://www.wynnsoft-solution.net/th/article/view/80/>
- Cloudflare คืออะไร จะเข้ามาช่วยยองค์กรของคุณได้อย่างไร? (2021, 12 7). Retrieved from nipa: <https://web.nipa.cloud/how-cloudflare-protect-your-corporate>
- Warakorn Pradiskul, P. M. (2021). Recommender System Using Collaborative Filtering A Case Study of Toyota Buzz Company Limited, 11–21.
- Thongchai Klayklueng, W. S.-n.-n. (2019). เทคนิคการคัดเลือกกลุ่มโหลดรายอาทิตย์สำหรับรองรับแผนการติดตั้งระบบผลิตไฟฟ้าพลังงานแสงอาทิตย์บนหลังคาเพื่อเพิ่มค่าదาระนี่ ประสิทธิภาพการใช้พลังงานไฟฟ้า Load Clustering Technique Application to PV Solar Rooftop Installation Planning for Improving Energy Efficiency, 134–148.
- ปราณี พึงวิชา, อ. ท. (2019). ศึกษาการแบ่งกลุ่มพฤติกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายลังคอมออนไลน์ Clustering of Jewellery Purchasing Behaviour through Social Network, 213–224.
- จักรินทร์ สันติรัตนภักดี, ศ. น. (2021). การออกแบบและพัฒนากระบวนการจำแนกข้อมูลเรียนรู้โดยสารสาขาวิชเพื่อติดแท็กปัญหาการให้บริการ, 77–90.
- วุฒิชัย, ว. (2013). การเปรียบเทียบวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างการเขียนกับโครงสร้างพยางค์ The Comparison of Thai Word Segmentation with Thai Writing Structures and Syllable Structures, 504–509.
- เจษ. (2021, 11 23). A Beginner's Guide to Scikit-learn . Retrieved from hashdork: <https://hashdork.com/th/scikit-learn/>

- mindphp. (2022, 6 23). การใช้งานต่างๆ ใน PyThaiNLP. Retrieved from mindphp:
<https://www.mindphp.com/บทความ/it-news/8778-การใช้งานต่างๆ ใน-pythainlp.html>
- JUNG. (2019). พื้นฐาน Python และ Numpy สำหรับ Deep Learning. Retrieved from kaggle:
<https://www.kaggle.com/code/ratthachat/python-numpy-deep-learning#Numpy>
- frevation. (2021, 10 8). Next js. Retrieved from frevation: <https://www.frevation.com/blog/web-development/next-js/>
- CloudHM. (2022, 7 6). บริการของ AWS มีจุดเด่นและนำไปใช้ประโยชน์ในด้านใดได้บ้าง. Retrieved from cloudhm: <https://blog.cloudhm.co.th/what-is-and-what-business-need-aws/>
- Chakrit. (2019, 5 29). similarity – ความเหมือนที่แตกต่าง. Retrieved from softnix:
<https://www.softnix.co.th/2019/05/29/similarity–ความเหมือนที่แตกต่าง/>
- DIGI. (2022, 11 17). รู้จัก Clustering Model คืออะไร. Retrieved from digi:
<https://digi.data.go.th/blog/what-is-clustering-model-and-example/>
- Chakrit. (2018, 9 6). ว่าด้วย-k-means-และการประยุกต์. Retrieved from softnix:
<https://www.softnix.co.th/2018/09/06/ว่าด้วย-k-means-และการประยุกต์/>
- CHAKRIT. (2019, 5 28). TF-IDF ทำงานยังไง. Retrieved from softnix:
<https://www.softnix.co.th/2019/05/28/tf-idf–ทำงานยังไง/>
- Paul. (2021, 3 4). K-Means Clustering with Elbow Method. Retrieved from medium:
<https://medium.com/kbtg-life/k-means-clustering-with-elbow-method-8d02b35aaa2e>
- Rungnapha, K. (2018, 1 31). Sequence Diagram แผนผังการทำงานแบบลำดับปฏิสัมพันธ์. Retrieved from glurgeek: <https://www.gurgeek.com/education/sequence-diagram/>
- Surapong, K. (2020, 1 7). PyThaiNLP คืออะไร Tutorial สอนใช้งาน PyThaiNLP Library NLP ภาษาไทย สำหรับ Python เป็องตัน – PyThaiNLP ep.1. Retrieved from bualabs:
<https://www.bualabs.com/archives/3234/what-is-pythainlp-tutorial-teach-basic-how-to-use-pythainlp-library-nlp-in-python-pythainlp-ep-1/>
- Panchart, M. (2021, 11 30). DATA รู้จัก pandas – Library อันดับ 1 สำหรับการทำ Data Analysis. Retrieved from skooldio: <https://blog.skooldio.com/what-is-pandas/>
- Pallop, C. (2017, 8 9). Next.js คืออะไร? Retrieved from medium:
<https://medium.com/hamcompe/next-js–คืออะไร-8fb36e68b0>
- PLC, V. M. (2022, 2 2). เข้าถึงและจัดการข้อมูลได้ง่ายๆ ด้วย MongoDB. Retrieved from proen:
<https://www.proen.cloud/en/blogs/mongodb/>
- Chai, P. (2015, 8 16). MongoDB คืออะไร? + สอนวิธีใช้งานเบ็องตัน. Retrieved from devahoy:
<https://devahoy.com/blog/2015/08/getting-started-with-mongodb>
- Natakorn, C. (2021, 6 3). FastAPI คืออะไร และการใช้งานเบ็องตัน. Retrieved from medium:
<https://natakornch.medium.com/fastapi–คืออะไร–และการใช้งานเบ็องตัน–4f2d0fd91bcd>
- TAeng Trirong, P. (2017, 12 25). Cross-Origin Resource Sharing (CORS) เป็นลิ๊งค์ Web Developer ต้องควรรู้. Retrieved from medium: <https://medium.com/nellika/cors–เป็นลิ๊งค์–web-developer–ต้องควรรู้–c906b1b47958>
- Supalerk, P. (2020, 3 17). เมื่อสาย DATA อยากจะกิน Pizza (โดยใช้ Jaccard Similarity และ Cosine Similarity). Retrieved from medium: <https://medium.com/data-cafe-thailand/เมื่อ>

- สาย-data-อ yak จะ กิน-pizza-โดยใช้-jaccard-similarity-และ-cosine-similarity-
f921fa4ab043
- Weerasak, T. (2017, 3 8). การหาจำนวน k ที่เหมาะสมที่สุดด้วยวิธี Elbow Method. Retrieved from medium: <https://medium.com/espressofx-notebook/การหาจำนวน-k-ที่เหมาะสมที่สุดด้วยวิธี-elbow-method-79b9a75f934>
- Patipan, P. (2020, 10 1). สถิติความสำคัญของข้อความด้วยเทคนิคการประมวลผลทางภาษา
เบื้องต้น: TF-IDF, Part 1. Retrieved from bigdata: <https://bigdata.go.th/big-data-101/tf-idf-1/>
- L, M. (2019, 2 1). NLP(Natural Language Processing) ศาสตร์(ไม่)ใหม่ ศาสตร์แห่งเจ้า: แยก
ประเภทอีเมลล์ด้วยพลังฟอร์ซ. Retrieved from medium: <https://medium.com/mmp-li/nlp-natural-language-processing-ศาสตร์-ไม่-ใหม่-ศาสตร์แห่งเจ้า-โดย-แยกประเภทอีเมลล์ด้วยพลังฟอร์ซ-66b8bdff2e42>

រាជធានីភ្នំពេញ

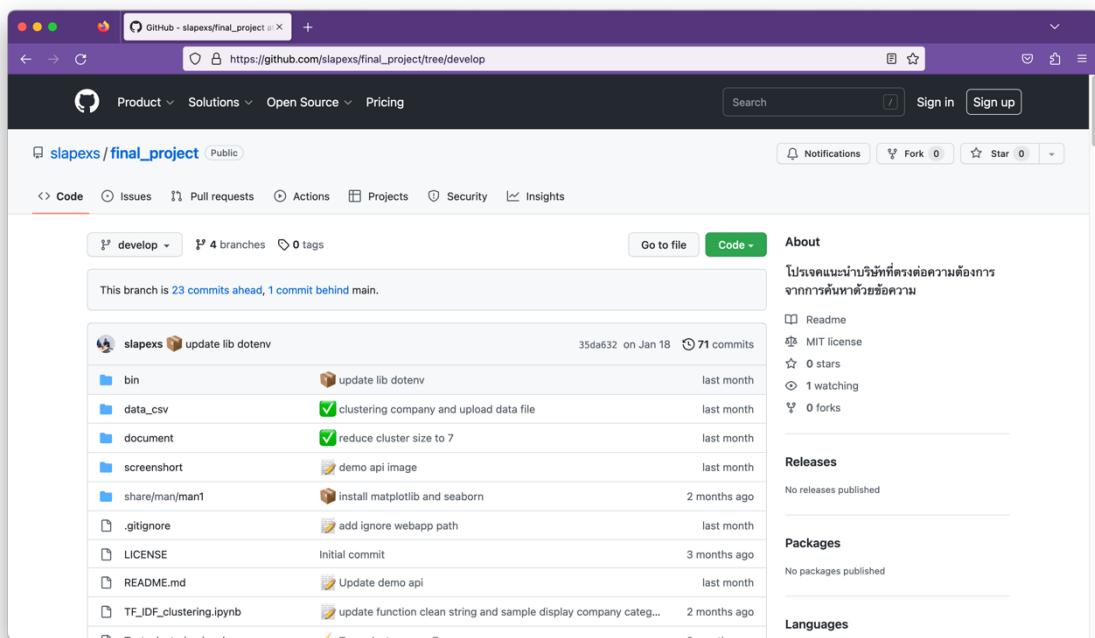
គ្រប់គ្រងការបង្កើតរំភ័យ

រាជធានីភ្នំពេញ គម្រោង ក្រសួង កម្មវិធាន

ការពេទ្យលេខាគម្ពស់បច្ចុប្បន្ន និងការពេទ្យលេខាគម្ពស់បច្ចុប្បន្ន ដែលបានរាយការណ៍នៅក្នុងក្រសួង

1. ការចែកចាយមូលដ្ឋានទៅតូអន្តេរដៃគ្រប់គ្រង់តាមផ្ទាល់ខ្លួន (K-Means)

1.1 ការទាញយក Repository ពី Github ដោយចូលរួមនៅក្នុង URL https://github.com/slapexs/final_project

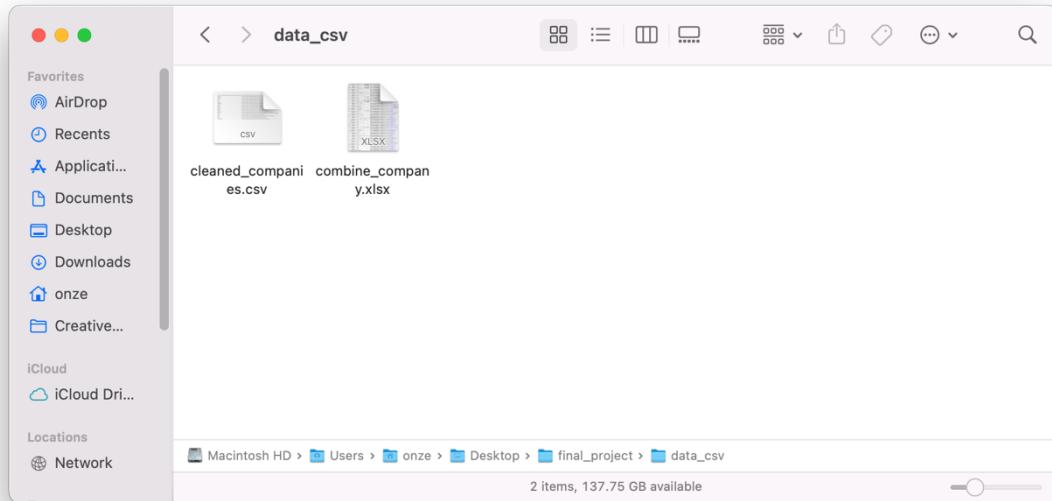


រាងទី 55 នៃ Repository នៃ Github

```
onze@Tinngrits-MacBook-Pro:~/desktop
git clone https://github.com/slapexs/final_project.git
Cloning into 'final_project'...
remote: Enumerating objects: 767, done.
remote: Counting objects: 100% (63/63), done.
remote: Compressing objects: 100% (49/49), done.
remote: Total 767 (delta 27), reused 46 (delta 14), pack-reused 704
Receiving objects: 100% (767/767), 17.56 MiB | 16.23 MiB/s, done.
Resolving deltas: 100% (369/369), done.
```

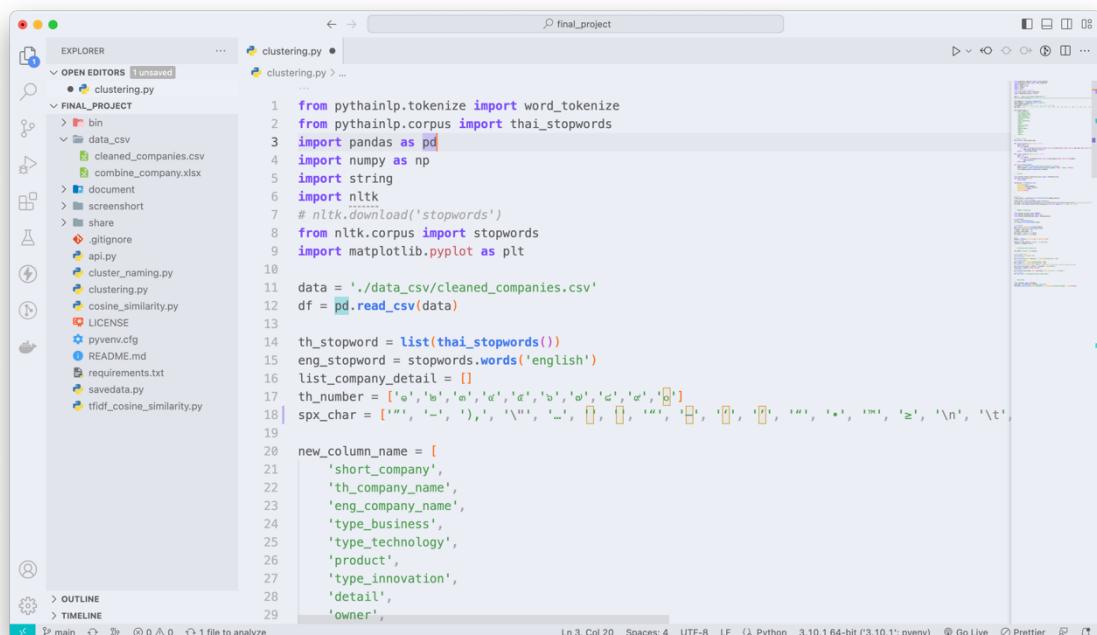
រាងទី 56 ការទាញយក Repository ដោយកំណាំការ 'git clone'

1.2 บันทึกไฟล์ข้อมูลบริษัทนามสกุล .CSV ในโฟลเดอร์ data_csv



ภาพที่ 57 แสดงการบันทึกไฟล์ข้อมูลบริษัทในโฟลเดอร์

1.3 เปิดไฟล์ clustering.py ในโปรแกรม Text editor



ภาพที่ 58 แสดงการเปิดไฟล์ clustering.py ในโปรแกรม Visual studio code

1.4 ทำการกำหนดจำนวนของกลุ่มข้อมูลที่ตัวแปร k

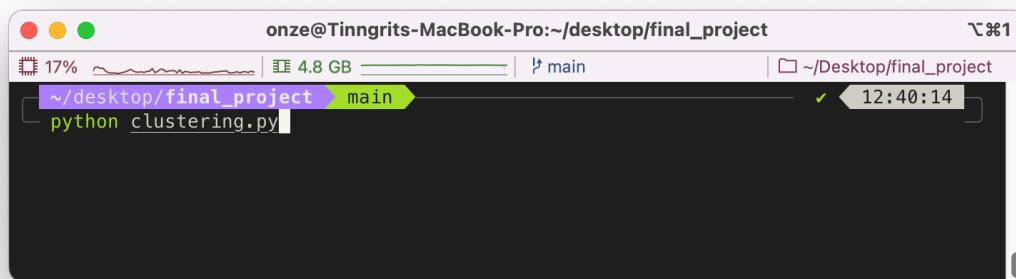
```

104    k = 7
105    kmeans = KMeans(n_clusters=k, random_state=1)
106    # # Fit model
107    kmeans.fit(df_tfidf[['x_value', 'y_value']])
108    clusters = kmeans.labels_

```

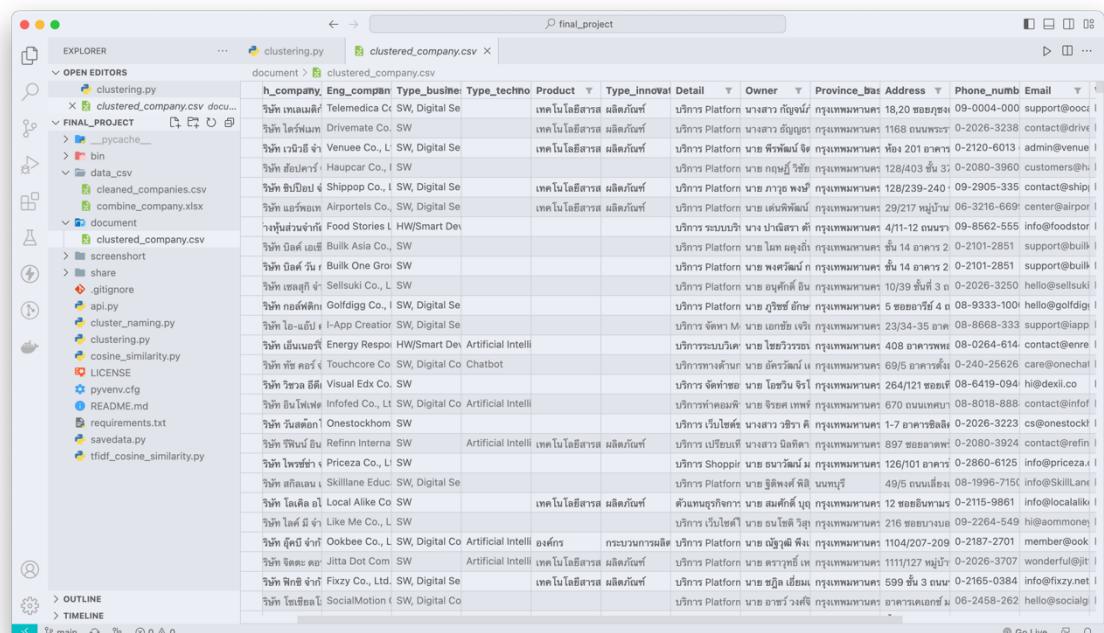
ภาพที่ 59 แสดงกำหนดจำนวนของกลุ่มที่ตัวแปร k

1.5 รันคำสั่งจัดกลุ่มข้อมูล



ภาพที่ 60 แสดงการรันคำสั่งจัดกลุ่มข้อมูลใน Terminal

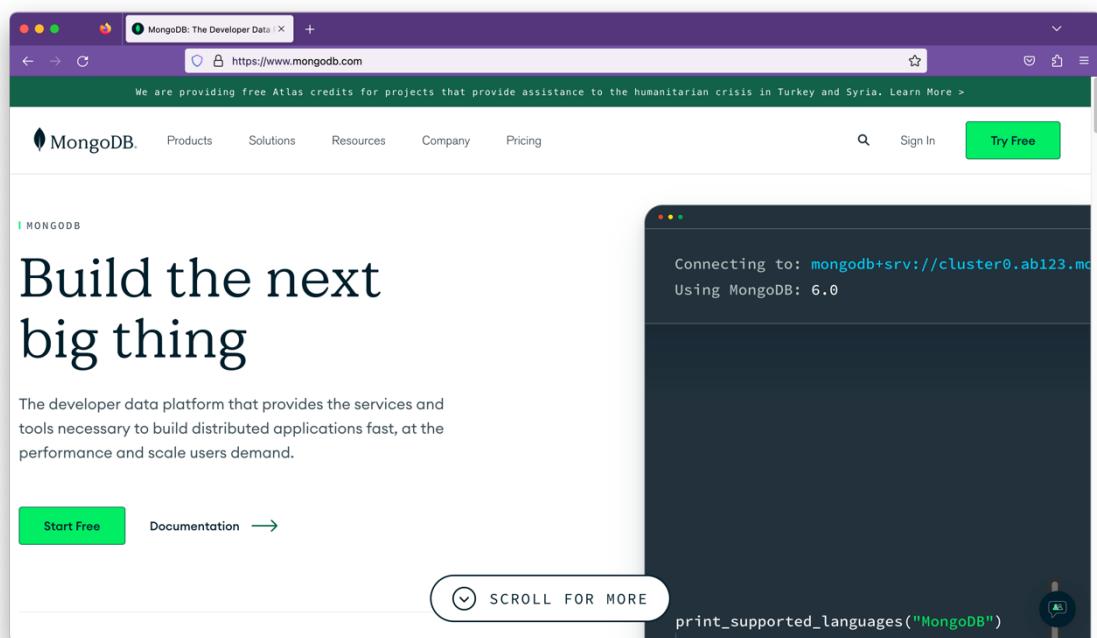
1.6 ได้ไฟล์ clustered_company.csv ในโฟลเดอร์ document ที่เป็นผลลัพธ์การจัดกลุ่มข้อมูล



ภาพที่ 61 แสดงไฟล์ clustered_company.csv

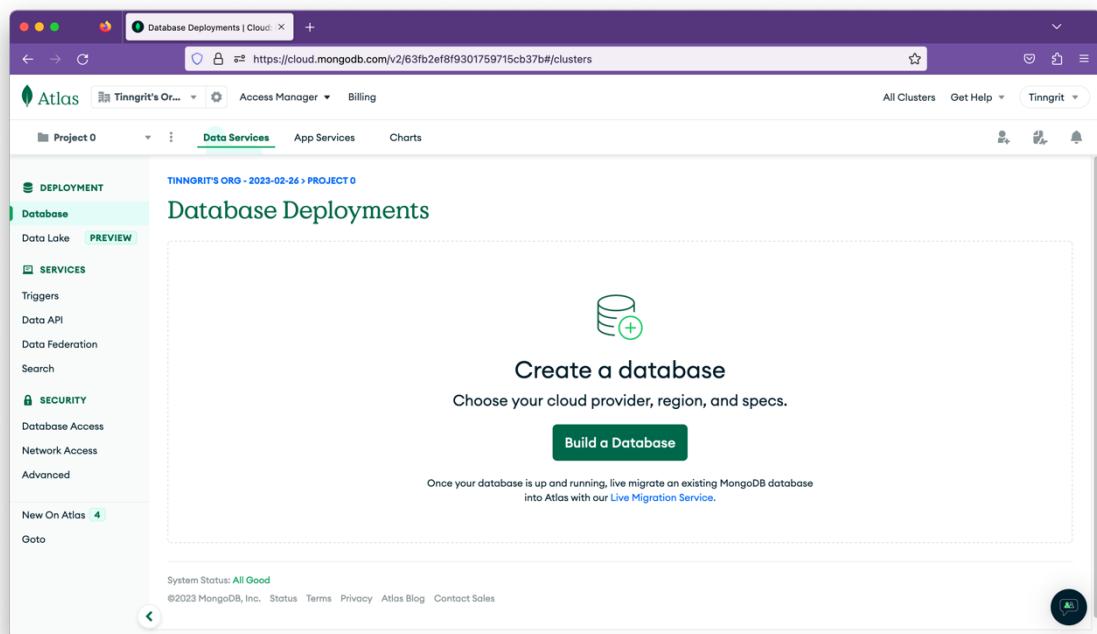
2. การสร้างคลัสเตอร์ (Cluster) ของฐานข้อมูล mongoDB บนเว็บไซต์

2.1 การใช้งาน Cloud MongoDB เข้าเว็บไซต์ <https://www.mongodb.com/> จากนั้นเข้าสู่ระบบ



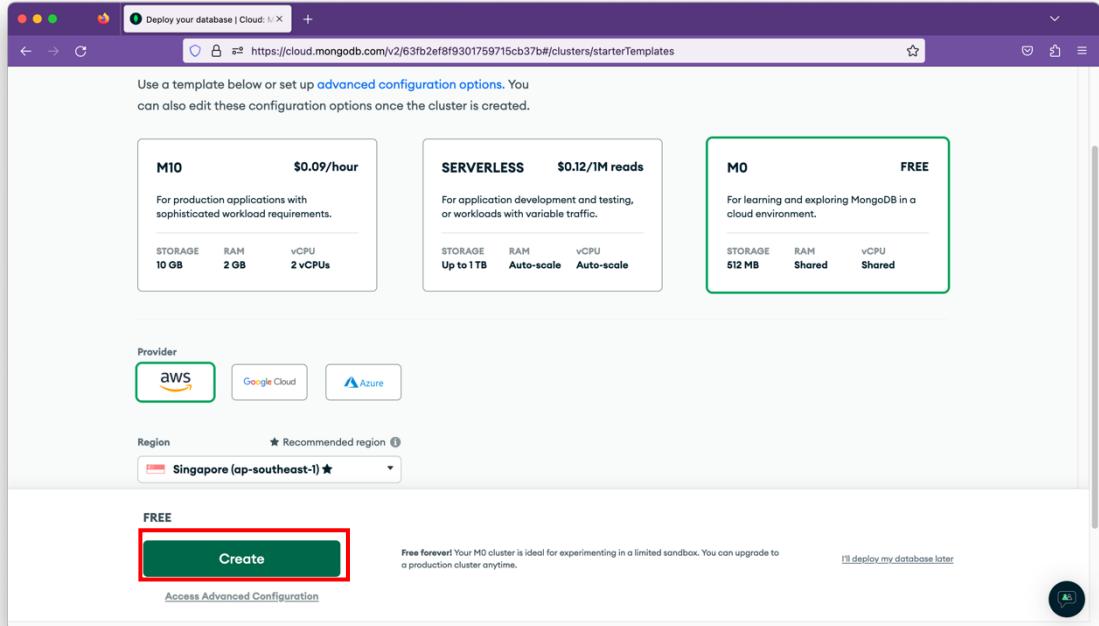
ภาพที่ 62 แสดงหน้าเว็บไซต์ MongoDB

2.2 สร้าง Cluster โดยการกดปุ่ม Build a Database



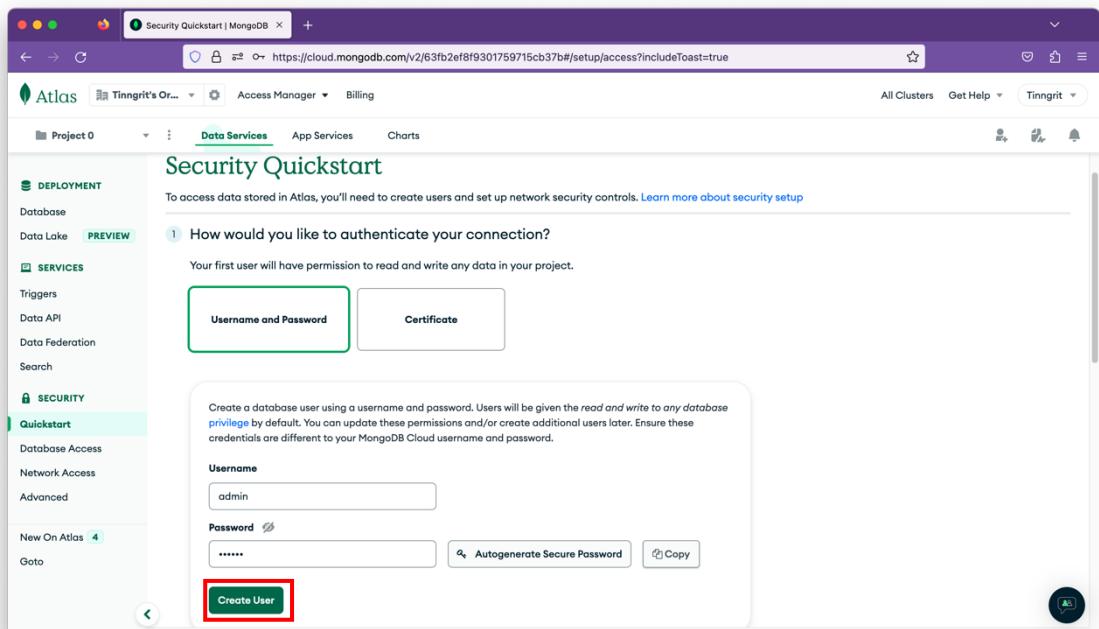
ภาพที่ 63 แสดงหน้าการจัดการ Cluster MongoDB

2.3 เลือกการตั้งค่าของ Cluster และตั้งชื่อ จากนั้นกดปุ่ม Create เพื่อสร้างฐานข้อมูล



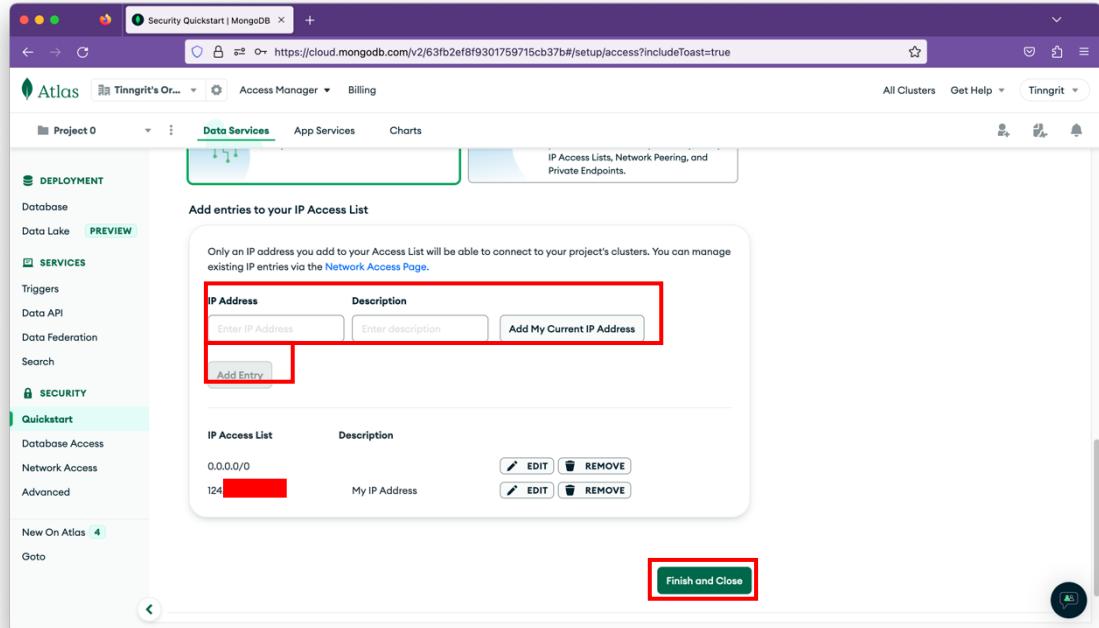
ภาพที่ 64 แสดงหน้าตั้งค่าและสร้าง Cluster MongoDB

2.4 สร้างบัญชีสำหรับใช้งานฐานข้อมูลตั้งค่า username และ password จากนั้นกดปุ่ม Create User



ภาพที่ 65 แสดงหน้าสร้างบัญชีสำหรับจัดการฐานข้อมูล

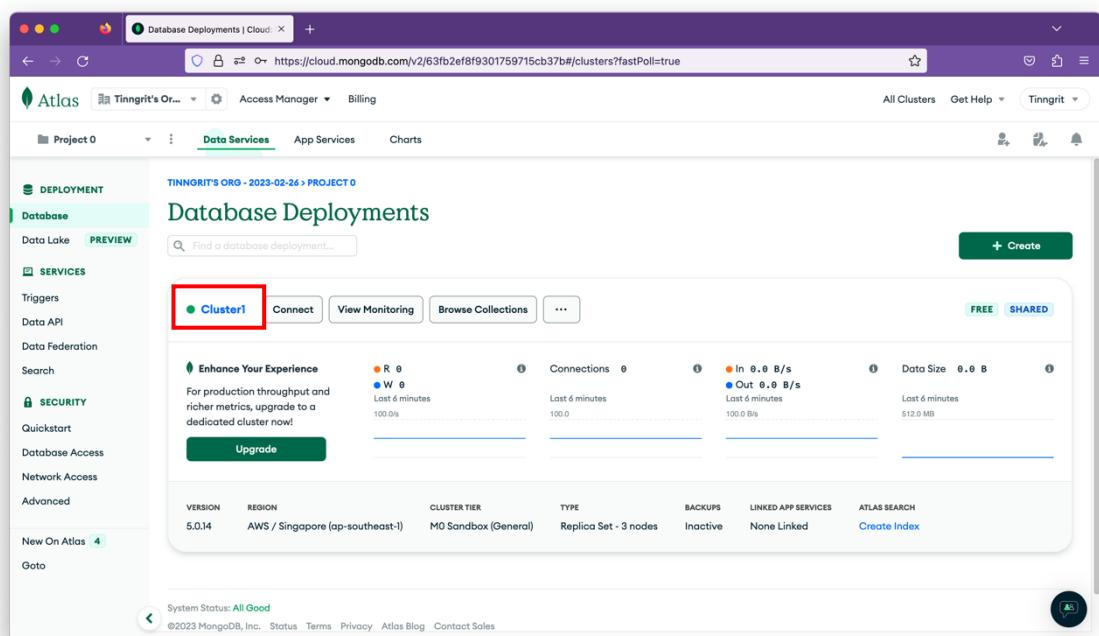
2.4 เพิ่มรายชื่อ IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้ (กรณีเป็น 0.0.0.0 หมายถึงทุก IP address สามารถเชื่อมต่อเข้ามาได้) จากนั้นกดปุ่ม Add Entry และปุ่ม Finish and Close



ภาพที่ 66 แสดงหน้าเพิ่ม IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้

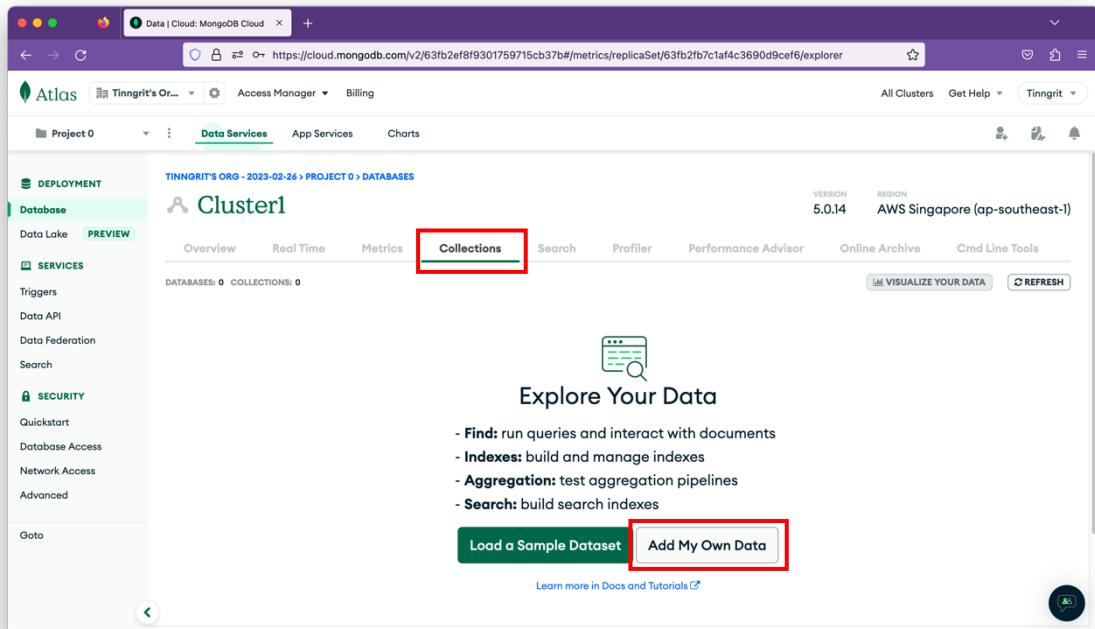
3. การสร้างฐานข้อมูลในโปรแกรม mongod (MongoDB)

3.1 สร้างฐานข้อมูลและ Collection กดที่ชื่อ Cluster



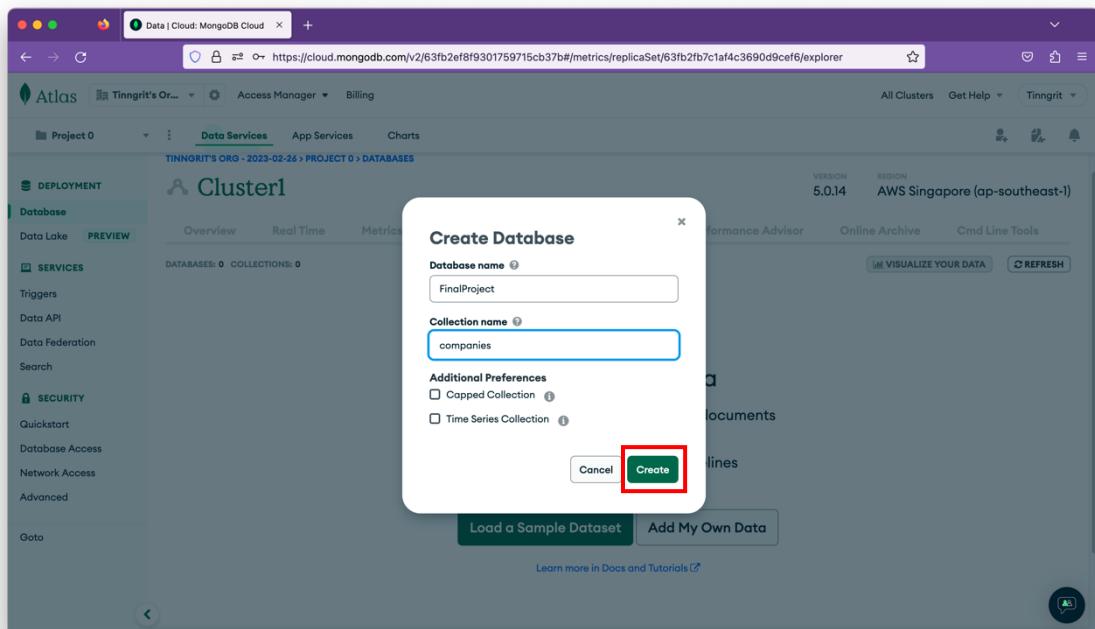
ภาพที่ 67 แสดงหน้าจัดการ Cluster MongoDB

3.2 กดที่ແນວເມນຸ Collections ແລະກດປຸ່ມ Add My Own Data ເພື່ອສ້າງຈຸານຂອ່ມູນ



ກາພທີ 68 ແສດນກາຮສ້າງຈຸານຂອ່ມູນ MongoDB

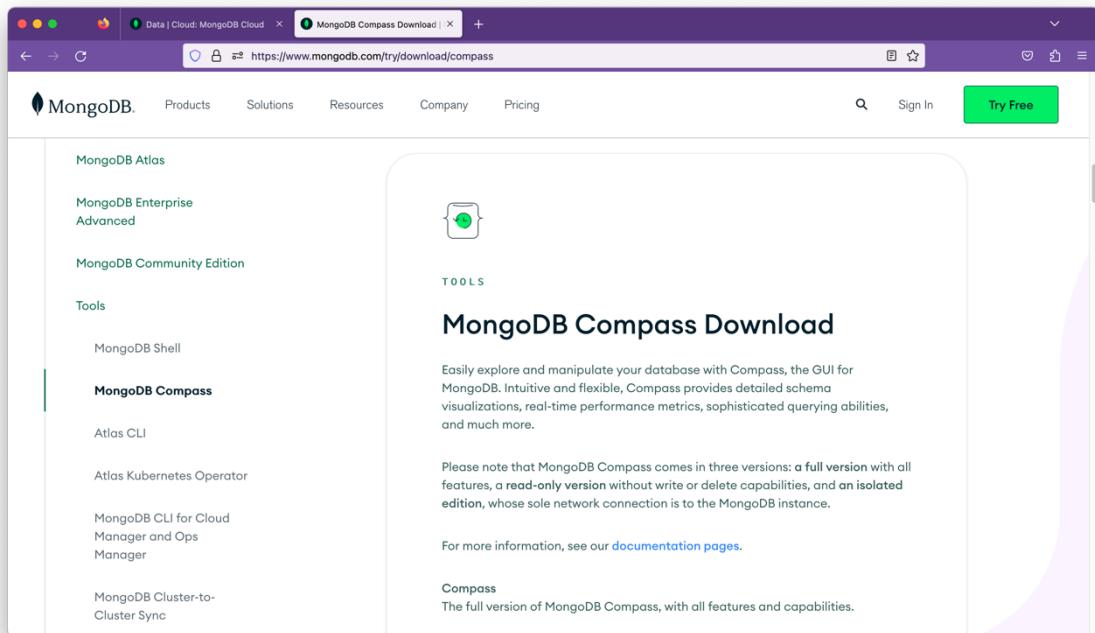
3.3 ກໍາທັນດີຂໍ້ອ່ຈຸານຂອ່ມູນແລະ ຂຶ້ນ Collection ທີ່ອໝື່ນຈຸານຂອ່ມູນ



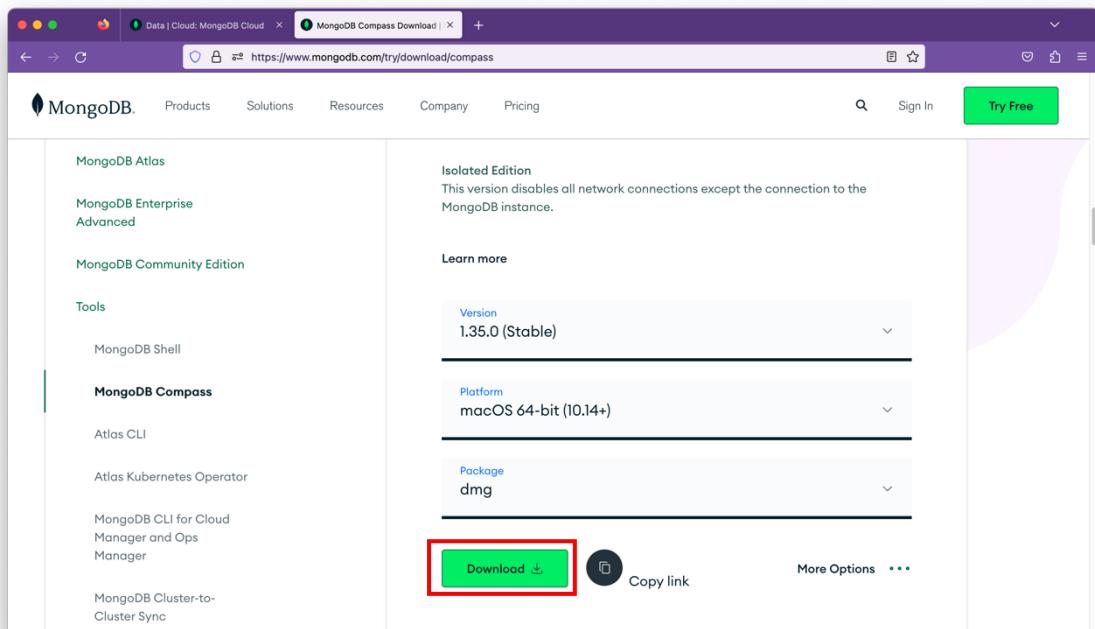
ກາພທີ 69 ແສດນໜ້າຕ່າງກາຮສ້າງຈຸານຂອ່ມູນແລະ Collection

3.3 ดาวน์โหลดโปรแกรม MongoDB compass ที่เว็บไซต์

<https://www.mongodb.com/try/download/compass> เลือก Version Platform และ Package ตามระบบปฏิบัติการที่ใช้ และกดปุ่ม Download



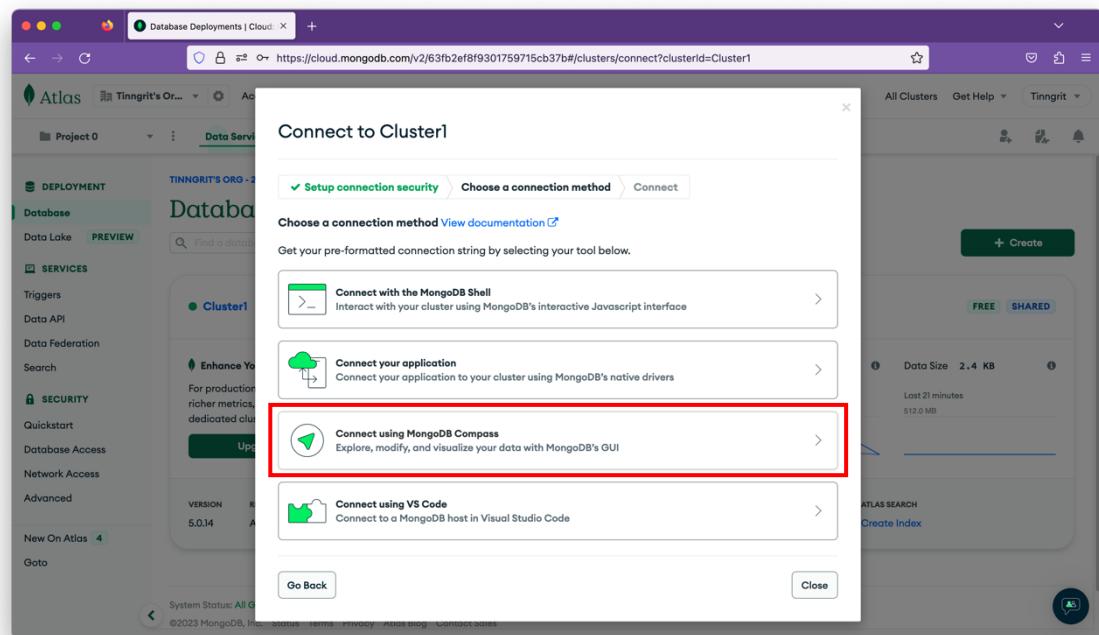
ภาพที่ 70 หน้าเว็บไซต์สำหรับโหลดโปรแกรม MongoDB compass



ภาพที่ 71 ตัวอย่างการเลือกตั้งค่าการดาวน์โหลดโปรแกรม MongoDB compass

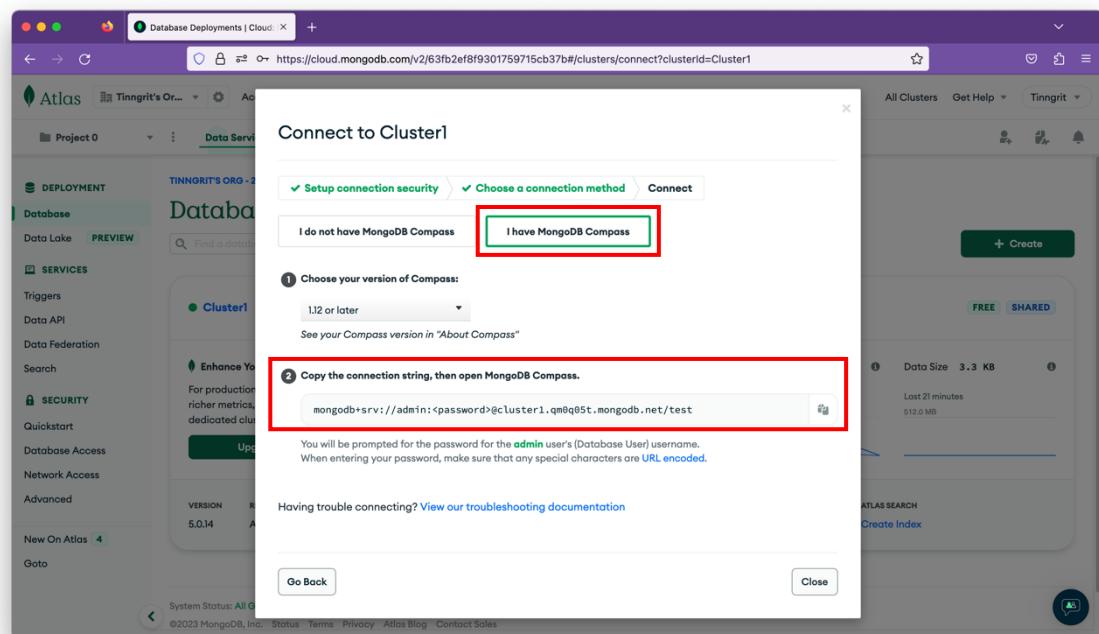
4. การนำข้อมูลเข้าสู่ฐานข้อมูลของโกลเด้น (MongoDB)

4.1 เปิดเว็บไซต์หน้าจัดการ Cluster กดปุ่ม Connect และกดปุ่ม Connect using MongoDB Compass



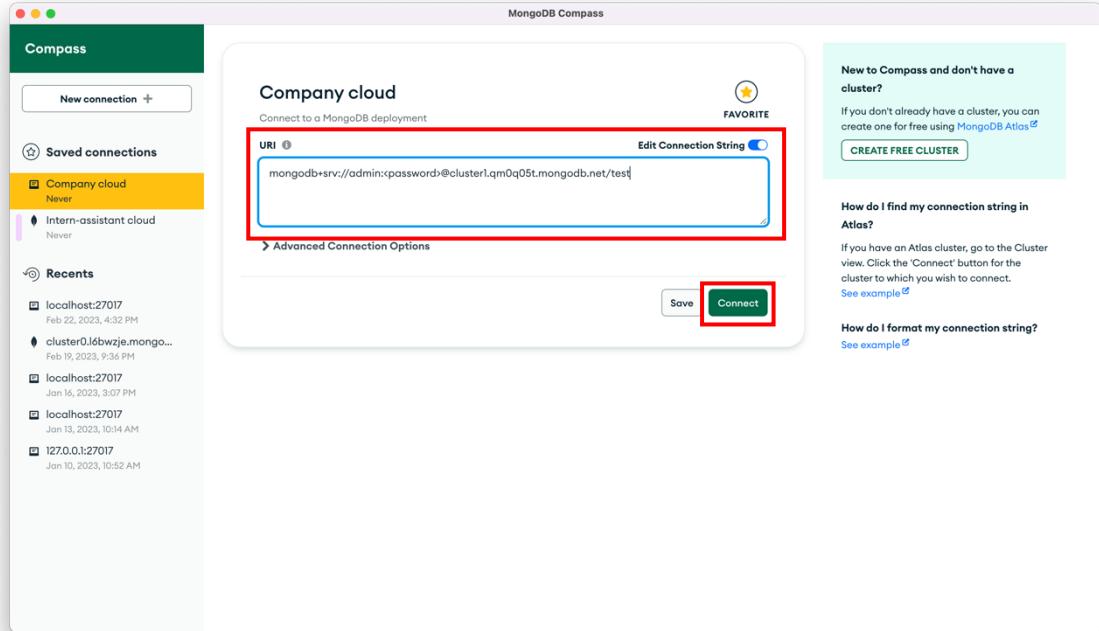
ภาพที่ 72 หน้าต่างเลือกเชื่อมต่อ กับ Cluster

4.2 กดปุ่ม I have MongoDB Compass และคัดลอก Connection string ไปข้อ 2



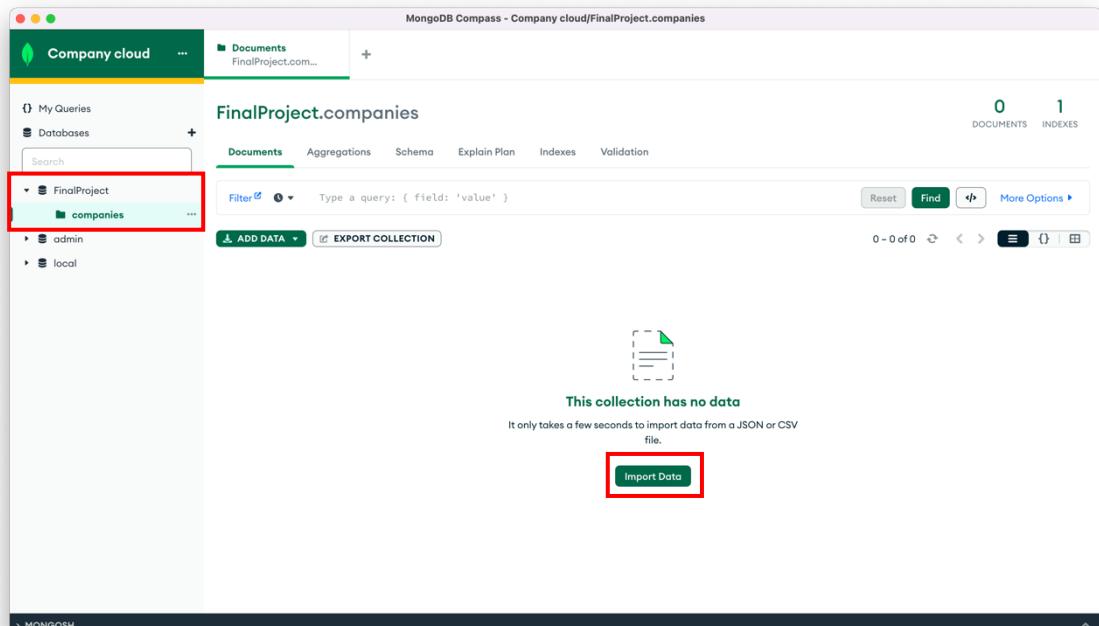
ภาพที่ 73 หน้าต่างขออนุญาตการเชื่อมต่อ Cluster กับ MongoDB compass

4.3 เปิดโปรแกรม MongoDB compass และวางแผนการเชื่อมต่อที่ช่อง URI แก้ไข username และ password ให้ตรงกับที่สร้างบัญชีไว้จากนั้นกดปุ่ม Connect



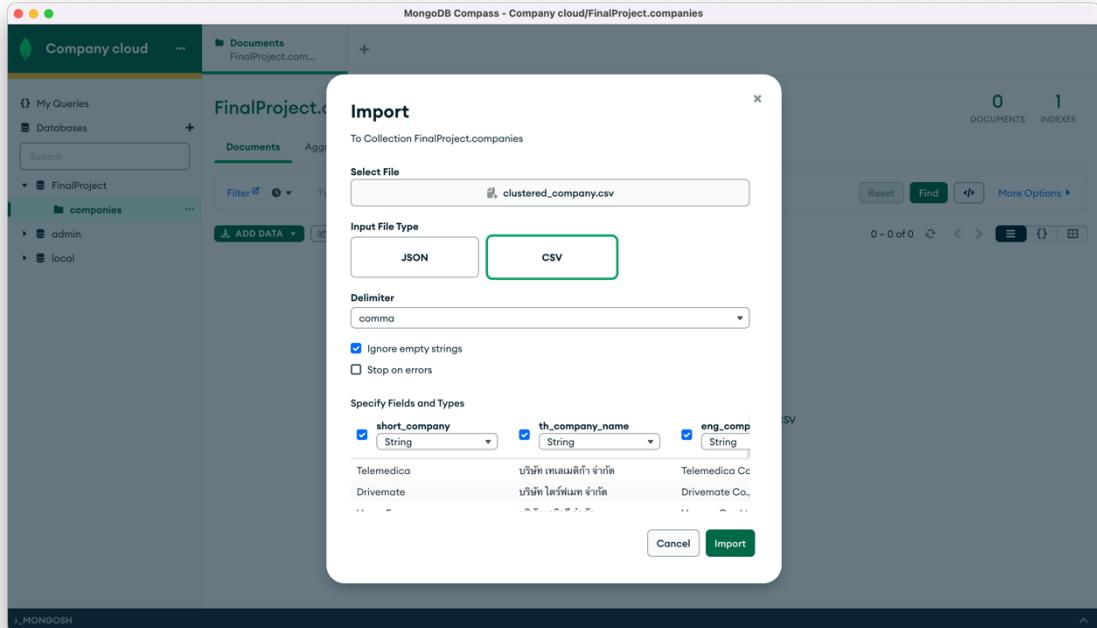
ภาพที่ 74 หน้าต่างโปรแกรม MongoDB compass สำหรับเชื่อมต่อ Cluster

4.4 เลือกฐานข้อมูลและ Collection ที่เมืองด้านซ้ายและกดปุ่ม Import Data



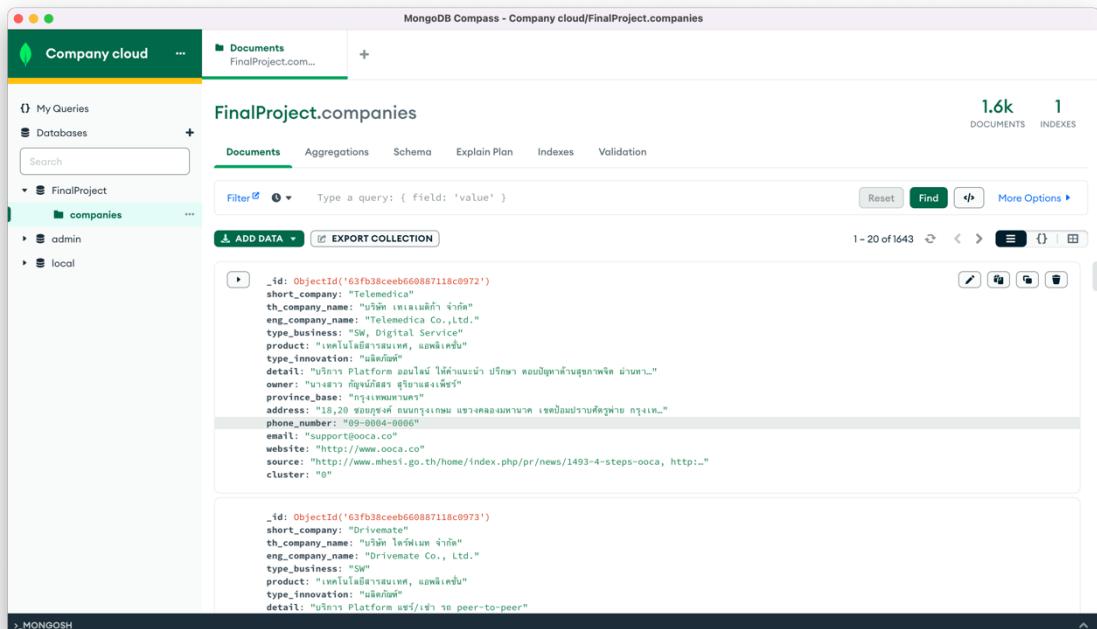
ภาพที่ 75 หน้าต่างโปรแกรมแสดงข้อมูลใน Collection

4.5 เลือกไฟล์ข้อมูลบริษัทที่จัดกลุ่มแล้ว กดปุ่ม CSV เพื่อ Import ข้อมูลแบบไฟล์นามสกุล CSV และกดปุ่ม Import และกดปุ่ม Done เพื่อเสร็จสิ้นกระบวนการ



ภาพที่ 76 หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv

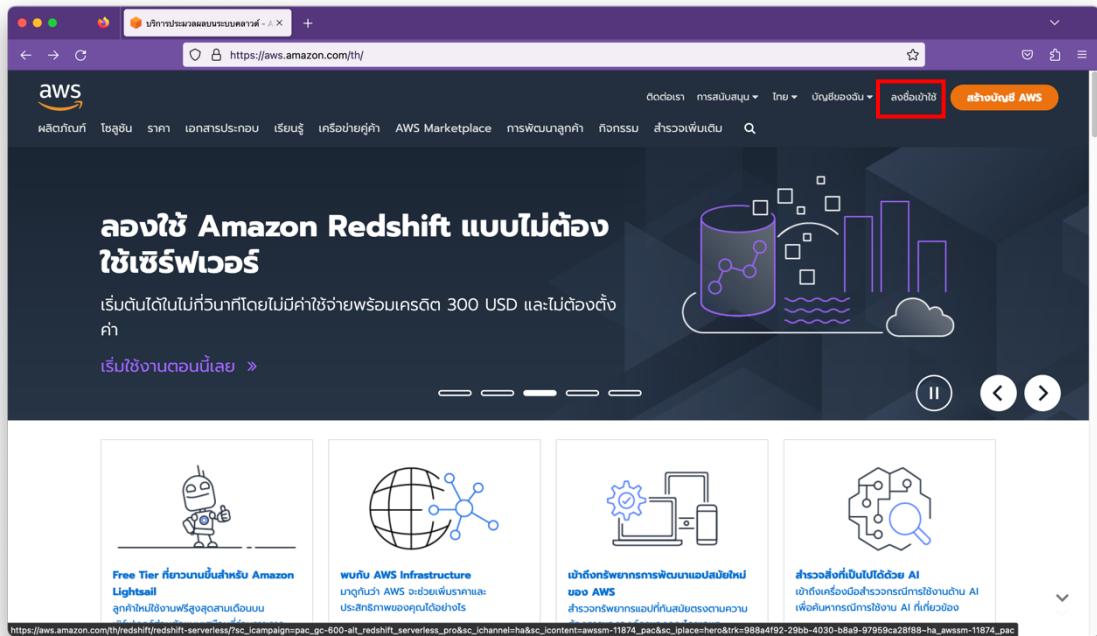
4.6 เมื่อ Import ข้อมูลสำเร็จจะได้ข้อมูลอยู่ใน Collection



ภาพที่ 77 หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass

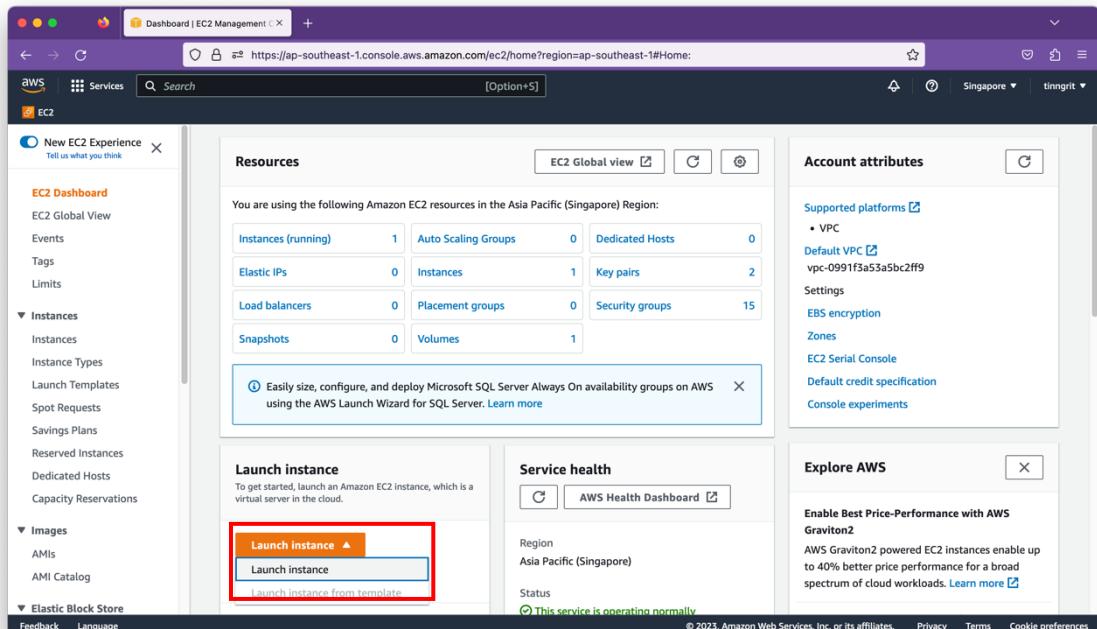
5. การใช้งาน Cloud computing ของ Amazon Web Services

5.1 เข้าไปยังเว็บไซต์ <https://aws.amazon.com/th/> เข้าสู่ระบบที่เมนู ลงชื่อเข้าใช้



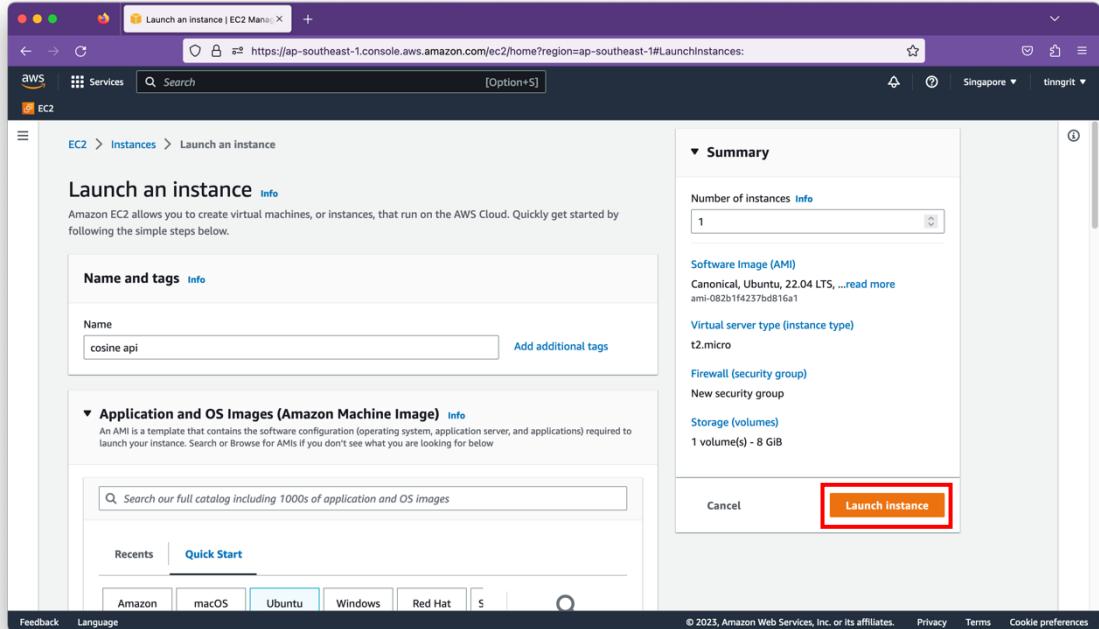
ภาพที่ 78 หน้าเว็บไซต์ Amazon Web Services

5.2 กดที่เมนู EC2 และกดปุ่ม Launch instance เพื่อสร้าง Instance ใหม่



ภาพที่ 79 หน้าแสดงการเลือกสร้าง Instance ใหม่

5.3 ตั้งค่าเครื่อง Instance ตามต้องการและกดปุ่ม Launch instance เพื่อทำการสร้าง



ภาพที่ 80 หน้าแสดงการตั้งค่า Instance

5.4 ทำการเชื่อมต่อไปยัง Instance ด้วยวิธี SSH โดยใช้ Terminal

```
ubuntu@ip-172-31-35-185:~$ cd desktop/aws
ubuntu@ip-172-31-35-185:~/desktop/aws$ ssh -i "finalproject.pem" ubuntu@ec2-52-221-246-234.ap-southeast-1.compute.amazonaws.com
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-1028-aws x86_64)

 * Documentation: https://help.ubuntu.com
 * Management: https://landscape.canonical.com
 * Support: https://ubuntu.com/advantage

 System information as of Sun Feb 26 12:32:02 UTC 2023

 System load: 0.24 Processes: 102
 Usage of /: 22.8% of 7.57GB Users logged in: 0
 Memory usage: 25% IPv4 address for eth0: 172.31.35.185
 Swap usage: 0%

 Expanded Security Maintenance for Applications is not enabled.

 24 updates can be applied immediately.
 18 of these updates are standard security updates.
 To see these additional updates run: apt list --upgradable

 Enable ESM Apps to receive additional future security updates.
 See https://ubuntu.com/esm or run: sudo pro status

 New release '22.04.2 LTS' available.
 Run 'do-release-upgrade' to upgrade to it.

 Last login: Sun Feb 26 12:31:43 2023 from 124.120.80.51
ubuntu@ip-172-31-35-185:~$
```

ภาพที่ 81 ตัวอย่างการเชื่อมต่อเข้าไปยัง Instance

6. การติดตั้งและใช้งาน Web API สำหรับคำนวณค่าความคล้ายคลึง (Cosine similarity)

6.1 ทำการ Clone project จาก Github ที่ลิงค์ https://github.com/slapexs/final_project.git

```
ubuntu@ip-172-31-35-185:~$ git clone https://github.com/slapexs/final_project.git
Cloning into 'final_project'...
remote: Enumerating objects: 769, done.
remote: Counting objects: 100% (65/65), done.
remote: Compressing objects: 100% (51/51), done.
remote: Total 769 (delta 27), reused 47 (delta 14), pack-reused 704
Receiving objects: 100% (769/769), 17.56 MiB | 14.54 MiB/s, done.
Resolving deltas: 100% (369/369), done.
ubuntu@ip-172-31-35-185:~$
```

ภาพที่ 82 การดาวน์โหลดโปรเจคจาก Github ด้วยคำสั่ง git clone มาอยู่ Instance และใช้คำสั่ง cd final_project เพื่อเข้าไปยังโฟลเดอร์

6.2 สร้างไฟล์ใหม่ในโฟลเดอร์ตั้งชื่อว่า .env และสร้างตัวแปรชื่อว่า MONGODB_USERNAME MONGODB_PASSWORD และ CLUSTER เพื่อให้เก็บข้อมูลเชื่อมต่อฐานข้อมูล

```
MONGODB_USERNAME=your_mongodb_username
MONGODB_PASSWORD=your_mongodb_password
CLUSTER=your_mongodb_cluster
```

ภาพที่ 83 สร้างไฟล์ใหม่ชื่อ .env และสร้างตัวแปรเพื่อกีบค่าเชื่อมต่อฐานข้อมูล

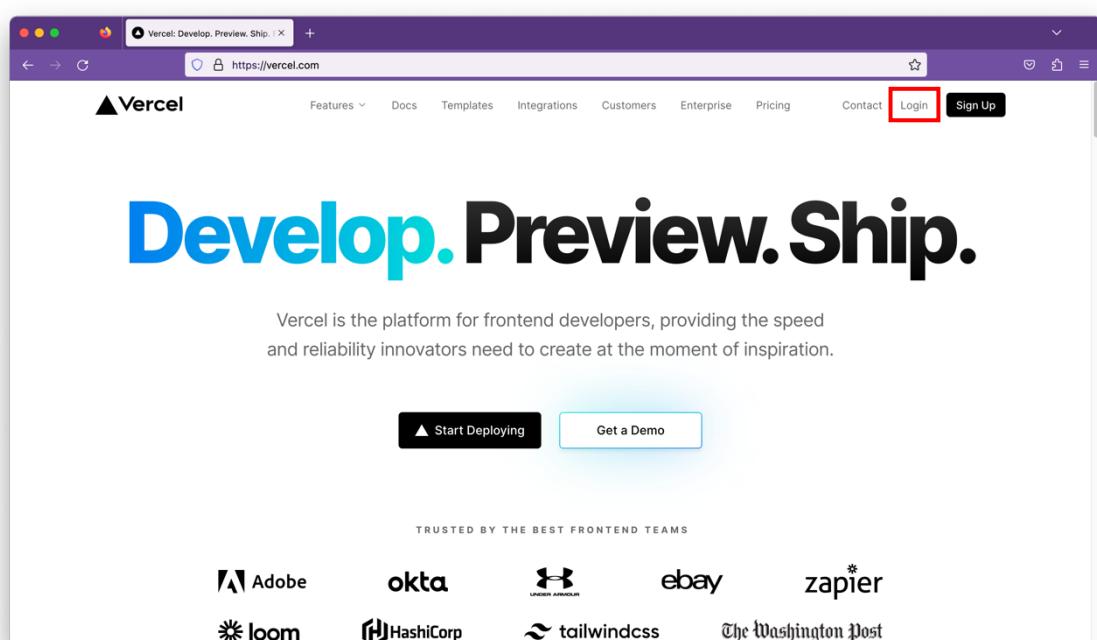
6.3 ใช้คำสั่ง pip3 install –r requirements.txt เพื่อทำการติดตั้ง library ที่จำเป็นและใช้คำสั่ง python3 –m uvicorn api:app เพื่อรัน Server Web API

```
ubuntu@ip-172-31-35-185:~/final_project$ python3 -m uvicorn api:app
INFO:     Started server process [20995]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)
```

ภาพที่ 84 ตัวอย่างการรัน Server Web API เพื่อคำนวณค่า Cosine similarity บน AWS

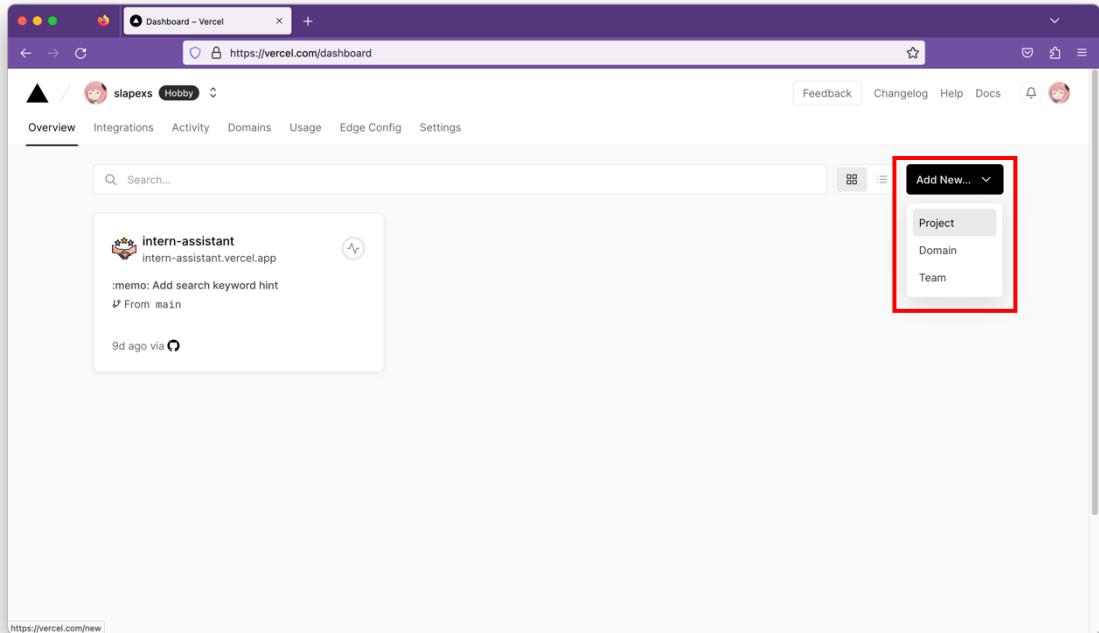
7. การติดตั้งและใช้งานเว็บแอปพลิเคชัน (Web application)

7.1 เข้าเว็บไซต์ <https://vercel.com> และเข้าสู่ระบบด้วยบัญชี Github



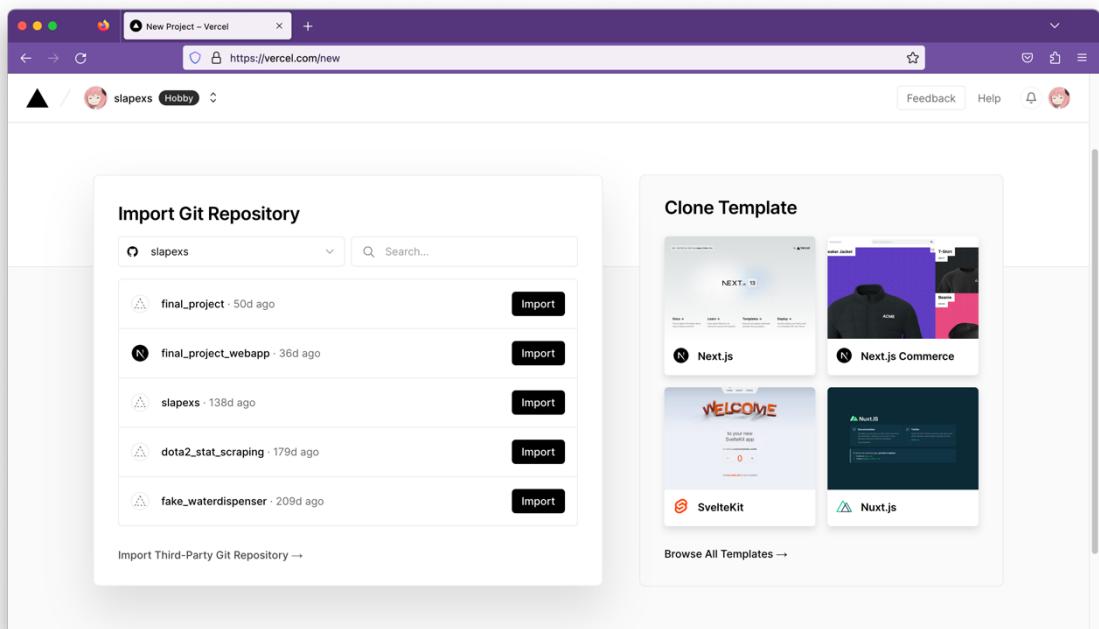
ภาพที่ 85 หน้าเว็บไซต์ Vercel

7.2 ทำการสร้างโปรเจคใหม่กดปุ่ม Add New และเลือก Project



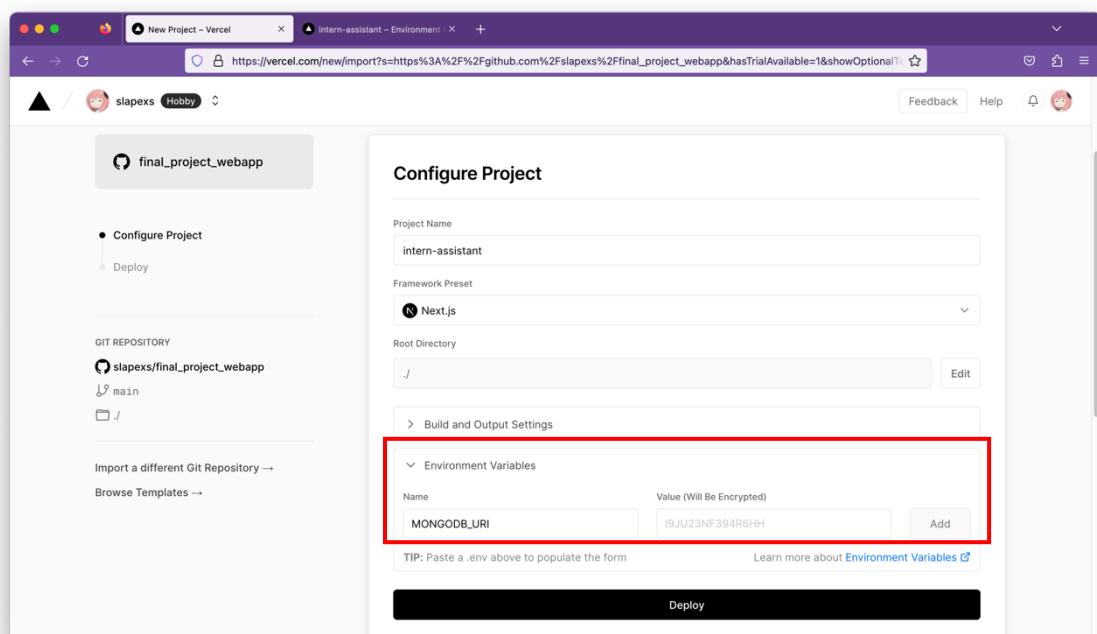
ภาพที่ 86 สร้างโปรเจคใหม่ใน Vercel

7.3 เลือก Repository ที่ต้องการจะ deploy และกดปุ่ม Import



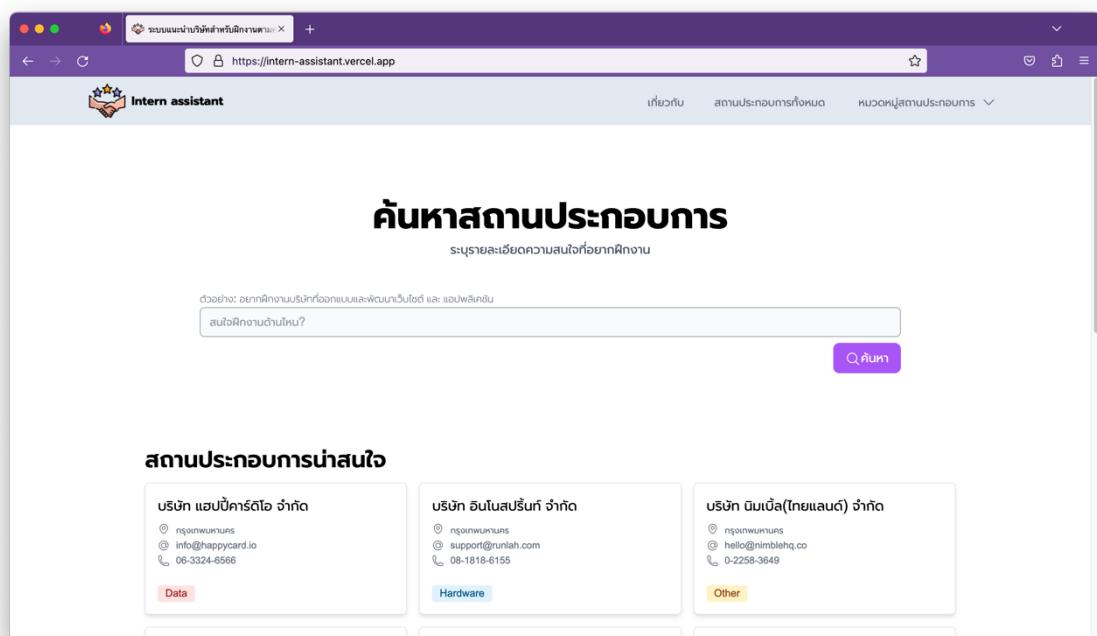
ภาพที่ 87 หน้าแสดงรายชื่อ Repository

7.4 ตั้งค่าโปรเจก และเพิ่มตัวแปร MONGODB_URI ในส่วนของ Environment Variables และใส่ค่าเป็น Connection string ของ MongoDB Atlas จากนั้นกดปุ่ม Deploy เพื่อทำการเผยแพร่สู่สาธารณะ



ภาพที่ 88 หน้าการตั้งค่าโปรเจกตอน Deploy

7.5 ตัวอย่างหน้าเว็บไซต์เมื่อ Deploy เรียบร้อย



ภาพที่ 89 ตัวอย่างหน้าเว็บไซต์

ภาคผนวก ข
คู่มือการใช้งาน

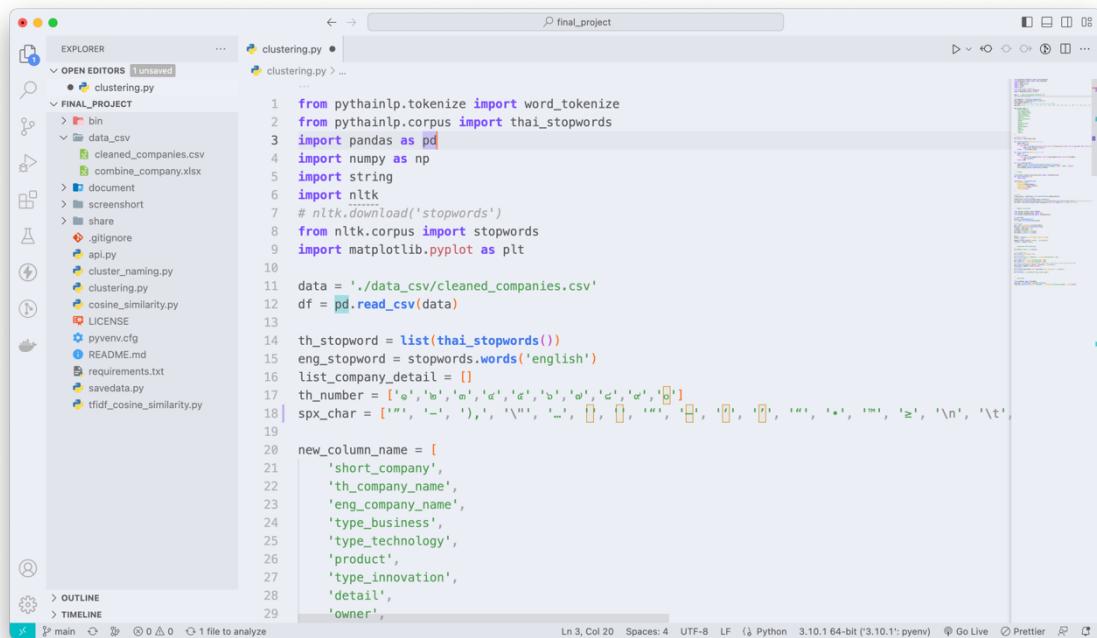
ภาคผนวก ข คู่มือการใช้งาน

ระบบแนะนำบริษัทสำหรับผู้ใช้งานตามความสนใจ ด้วยเทคโนโลยีปัญญาประดิษฐ์ สามารถแบ่งได้ 2 ส่วนดังนี้

1. ผู้ดูแลระบบ

1.1 การจัดกลุ่มข้อมูล

1 เปิดไฟล์ clustering.py ในโปรแกรม Text editor



```

OPEN EDITORS 1 unsaved
clustering.py
clustering.py > ...
1 from pythainlp.tokenize import word_tokenize
2 from pythainlp.corpus import thai_stopwords
3 import pandas as pd
4 import numpy as np
5 import string
6 import nltk
7 # nltk.download('stopwords')
8 from nltk.corpus import stopwords
9 import matplotlib.pyplot as plt
10
11 data = './data_csv/cleaned_companies.csv'
12 df = pd.read_csv(data)
13
14 th_stopword = list(thai_stopwords())
15 eng_stopword = stopwords.words('english')
16 list_company_detail = []
17 th_number = [\'\u0e31\', \'\\u0e30\', \'\\u0e32\', \'\\u0e33\', \'\\u0e34\', \'\\u0e35\', \'\\u0e36\', \'\\u0e37\', \'\\u0e38\', \'\\u0e39\', \'\\u0e3a\', \'\\u0e3b\', \'\\u0e3c\', \'\\u0e3d\', \'\\u0e3e\', \'\\u0e3f\', \'\\u0e31\']
18 spx_char = [\'\\n\', \'\\t\', \'\\r\', \'\\u0e31\', \'\\u0e30\', \'\\u0e32\', \'\\u0e33\', \'\\u0e34\', \'\\u0e35\', \'\\u0e36\', \'\\u0e37\', \'\\u0e38\', \'\\u0e39\', \'\\u0e3a\', \'\\u0e3b\', \'\\u0e3c\', \'\\u0e3d\', \'\\u0e3e\', \'\\u0e3f\', \'\\u0e31\']
19
20 new_column_name = [
21     'short_company',
22     'th_company_name',
23     'eng_company_name',
24     'type_business',
25     'type_technology',
26     'product',
27     'type_innovation',
28     'detail',
29     'owner',
30 ]

```

ภาพที่ 90 แสดงการเปิดไฟล์ clustering.py ในโปรแกรม Visual studio code

2. ทำการกำหนดจำนวนของกลุ่มข้อมูลที่ตัวแปร k

```

104 k = 7
105 kmeans = KMeans(n_clusters=k, random_state=1)
106 # # Fit model
107 kmeans.fit(df_tfidf[['x_value', 'y_value']])
108 clusters = kmeans.labels_

```

ภาพที่ 91 แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k

3. รันคำสั่งจัดกลุ่มข้อมูล

```
enze@Tinngrits-MacBook-Pro:~/desktop/final_project
[1] 17% 4.8 GB main ~ ~/Desktop/final_project
~/desktop/final_project main
python clustering.py
12:40:14
```

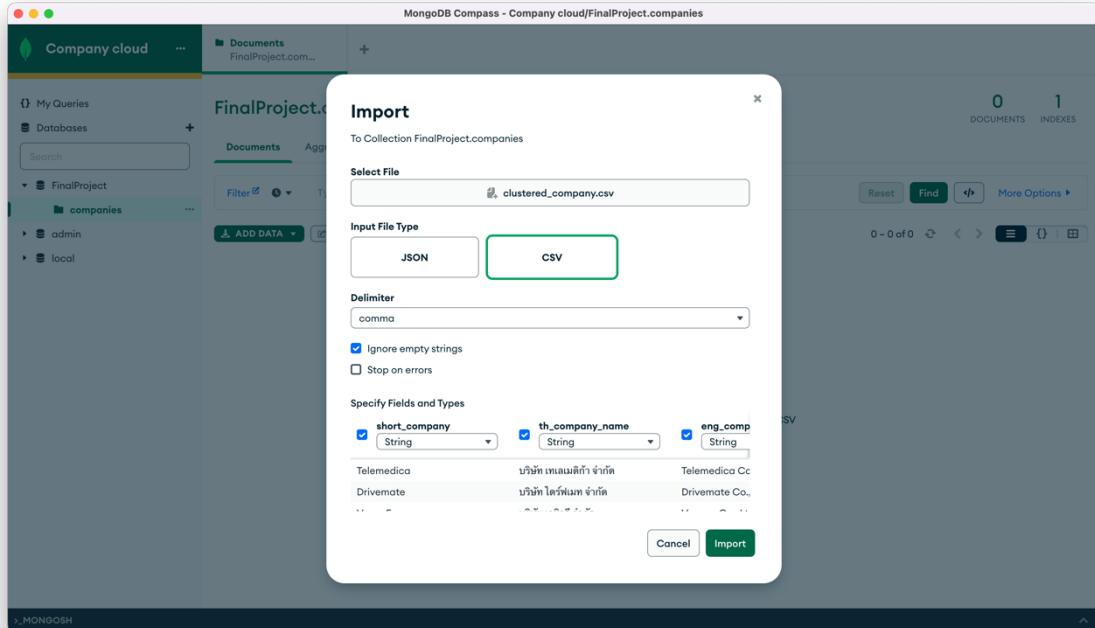
ภาพที่ 92 แสดงการรันคำสั่งจัดกลุ่มข้อมูลใน Terminal

4. ได้ไฟล์ clustered_company.csv ในโฟลเดอร์ document ที่เป็นผลลัพธ์การจัดกลุ่มข้อมูล

File	Type	Content Summary
clustering.py	Python Script	Script for clustering companies.
clustered_company.csv	CSV File	Clustering results for companies.

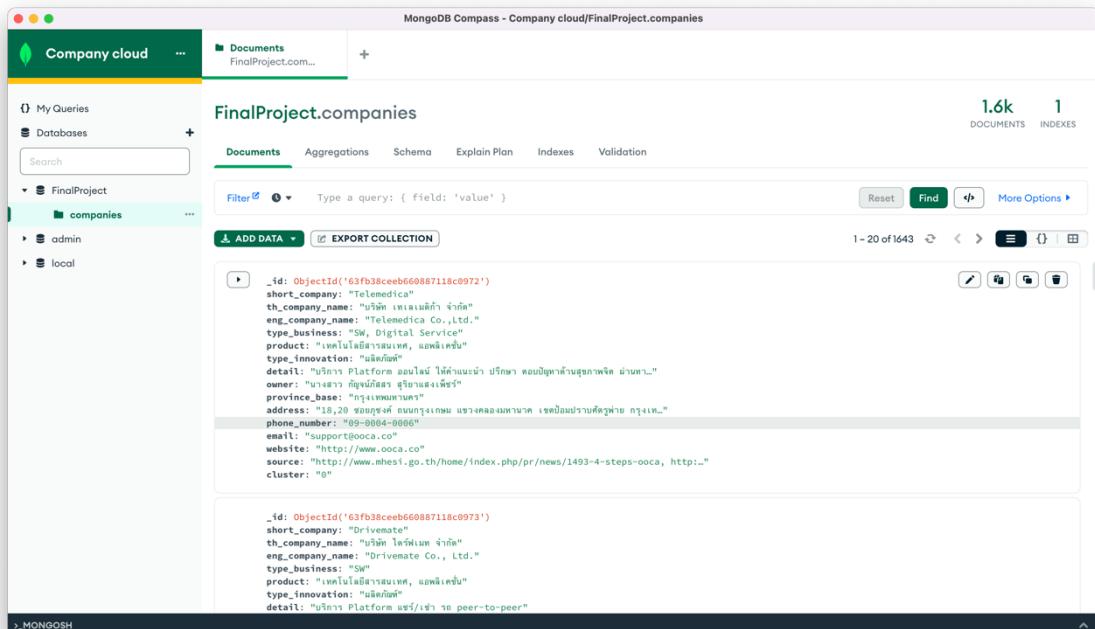
ภาพที่ 93 แสดงไฟล์ clustered_company.csv

5. เลือกไฟล์ข้อมูลบริษัทที่จัดลุ่มแล้ว กดปุ่ม CSV เพื่อ Import ข้อมูลแบบไฟล์นามสกุล csv และกดปุ่ม Import และกดปุ่ม Done เพื่อเสร็จสิ้นกระบวนการ



ภาพที่ 94 หน้าต่าง Import ข้อมูลนามสกุลไฟล์ CSV

6 เมื่อ Import ข้อมูลสำเร็จจะได้ข้อมูลอยู่ใน Collection

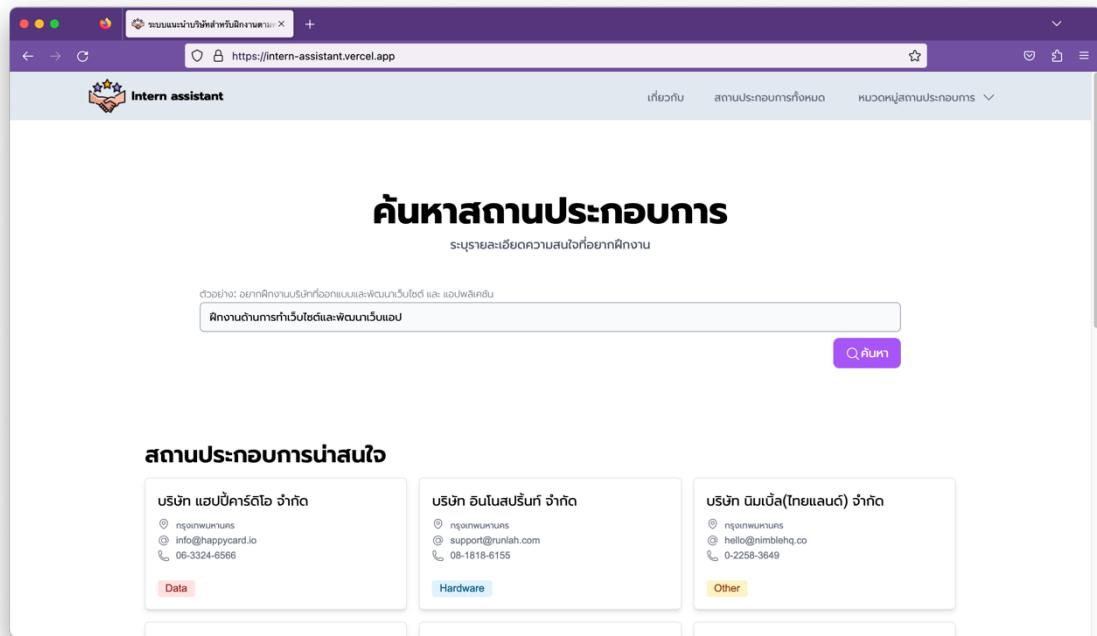


ภาพที่ 95 หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass

2. ផ្តើមឱ្យងារ

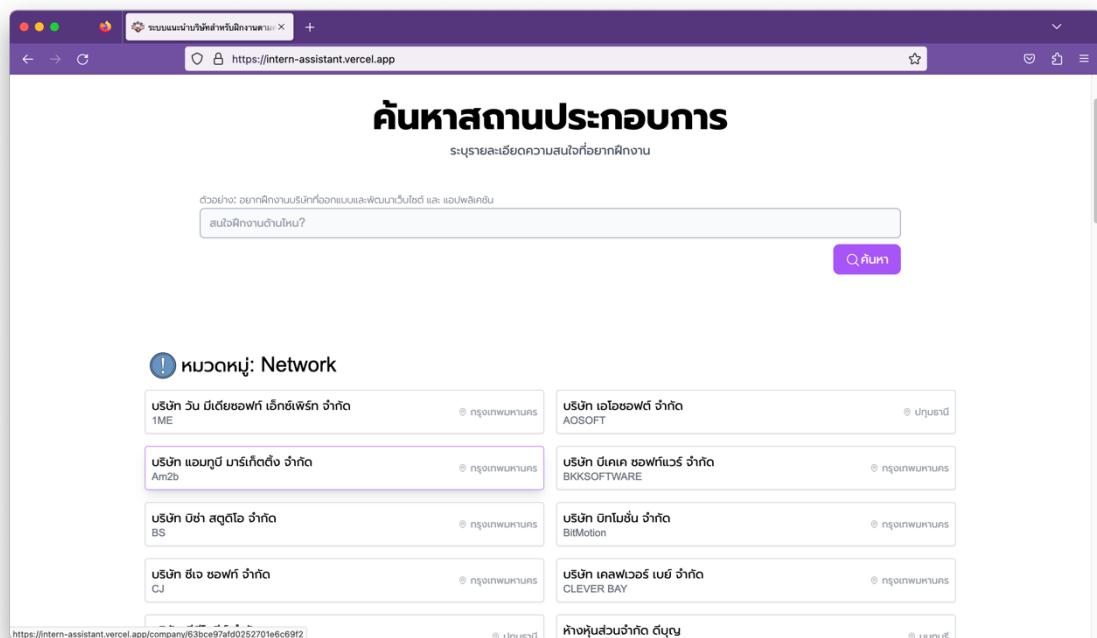
2.1 មេន្ទីរចូលទៅកាន់ព័ត៌មានប្រើប្រាស់នៃការគាំទ្រក្នុងការប្រើប្រាស់បច្ចុប្បន្ន។

ស្មើនឹងរូបរបៀបបង្ហាញខាងក្រោម



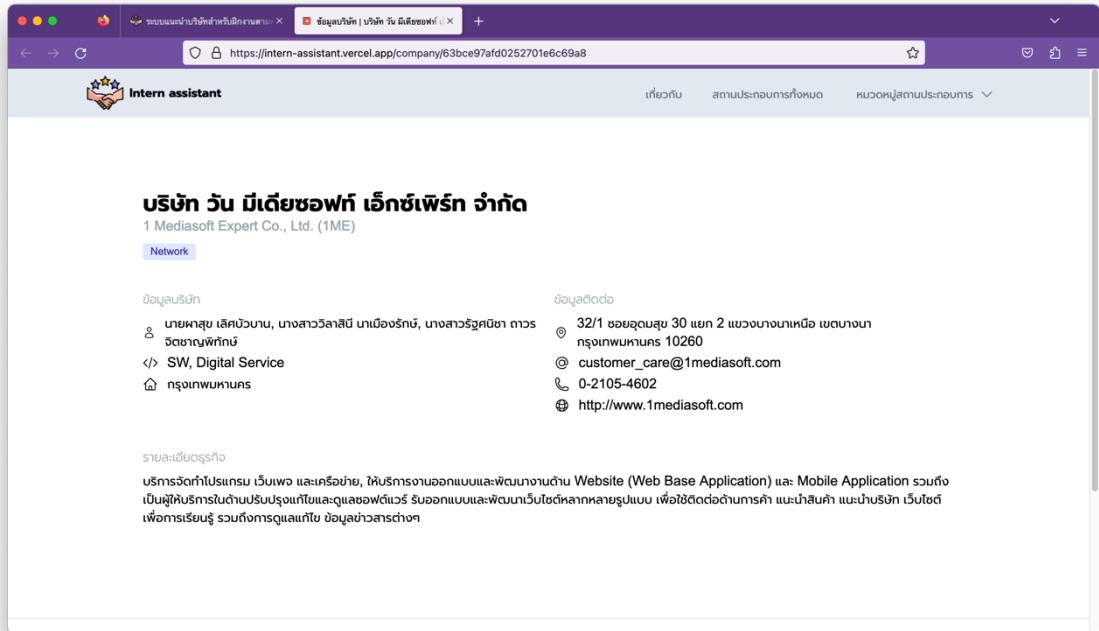
រាជធានី ៩៦ អន្តោះក្នុងព័ត៌មាន intern-assistant.vercel.app

2.2 មេន្ទីរចូលទៅកាន់ព័ត៌មានប្រើប្រាស់បច្ចុប្បន្ន។



រាជធានី ៩៧ អន្តោះក្នុងព័ត៌មាន intern-assistant.vercel.app

2.3 ផ្នែកទិន្នន័យសាមារណ៍ទូទៅមូលបរិមាណដើម្បីការត្រួតពិនិត្យ



រាយការណ៍ទូទៅមូលបរិមាណ

ประวัติผู้ศึกษา



ชื่อ-นามสกุล : นายทินกรถ สิงห์แกร้ว

รหัสนักศึกษา : 64342205007-7

วันเดือนปีเกิด : 22 สิงหาคม พ.ศ. 2541

ที่อยู่ปัจจุบัน : 209 ม.5 ต.ร้องกวาง อ.ร้องกวาง จ.แพร่ 54140

E-mail : tinngrit@outlook.com

ประวัติการศึกษา

พ.ศ. 2554 – พ.ศ. 2557 : สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้น
โรงเรียนร้องกวางอนุสรณ์ จ.แพร่

พ.ศ. 2557 – พ.ศ. 2560 : สำเร็จการศึกษามัธยมศึกษาตอนปลาย
โรงเรียนร้องกวางอนุสรณ์ จ.แพร่

พ.ศ. 2560 – พ.ศ. 2563 : สำเร็จการศึกษาระดับประกาศนียบัตรวิชาชีพชั้นสูง
สาขาวิชาเทคโนโลยีคอมพิวเตอร์
สาขางานคอมพิวเตอร์ซอฟต์แวร์
วิทยาลัยเทคนิคแพร่

พ.ศ. 2564 – ปัจจุบัน : กำลังศึกษาระดับปริญญาตรี สาขาวิทยาศาสตร์
หลักสูตรวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน