



## โครงการทางวิทยาการคอมพิวเตอร์ระดับปริญญาตรี

เรื่อง

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่ม  
เคลื่อนที่ (K-Means)

Computer Internship Recommendation System With  
K-Means Clustering

โดย

นายทินกฤต สิงห์แก้ว

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน  
ปีการศึกษา 2565



ใบรับรองโครงการงานวิชาการคณิตศาสตร์  
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน<sup>๑</sup>  
วิทยาศาสตรบัณฑิต (วิทยาการคณิตศาสตร์)

เรื่อง ระบบแนะนำสถานที่ฝึกงานด้านคณิตศาสตร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means)  
Computer Internship Recommendation System With K-Means Clustering  
โดย นายพินกฤต สิงห์แก้ว

คณะกรรมการพิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษา..... วันที่...../...../.....  
( ผู้ช่วยศาสตราจารย์ ดร.นงนุช เกตุชัย )

อาจารย์ที่ปรึกษาร่วม..... วันที่...../...../.....  
( อาจารย์วรวิทย์ ผันคำอ้าย )

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

ประธานหลักสูตร..... วันที่...../...../.....  
( อาจารย์วรวิทย์ ผันคำอ้าย )

# โครงการทางวิทยาการคอมพิวเตอร์ระดับปริญญาตรี

เรื่อง

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่ม  
เคลื่อนที่ (K-Means)

Computer Internship Recommendation System With K-Means Clustering

โดย

นายทินกฤต สิงห์แก้ว

คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร  
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน  
เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรบัณฑิต สาขาวิทยาศาสตร์  
พ.ศ. 2565

## บทคัดย่อ

<b>ชื่อโครงการ</b>	: ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means)
<b>อาจารย์ที่ปรึกษา</b>	: ผู้ช่วยศาสตราจารย์ ดร. นงนุช เกตุย
<b>อาจารย์ที่ปรึกษาร่วม</b>	: อาจารย์วรรธน์ ผันคำอักษย
<b>สาขาวิชา</b>	: วิทยาศาสตร์
<b>หลักสูตร</b>	: วิทยาการคอมพิวเตอร์
<b>ปีการศึกษา</b>	: 2565

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means) (Computer Internship Recommendation System With K-Means Clustering) เป็นระบบที่ช่วยแนะนำสถานประกอบการสำหรับการฝึกงานของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าสนใจในรูปแบบของเว็บแอปพลิเคชันที่พัฒนาโดยใช้ Next.js เว็บเฟรมเวิร์ค (Web Framework) สำหรับพัฒนาเว็บแอปพลิเคชัน โดยให้นักศึกษาระบุรายละเอียดความสนใจตามรูปแบบคร่าวๆ หรือรูปแบบของงานเพื่อนำมาวิเคราะห์หาความคล้ายคลึงกับข้อมูลสถานประกอบการที่มีอยู่ในฐานข้อมูลที่ผ่านกระบวนการวิเคราะห์ความคล้ายคลึงของความสนใจที่คนหา จากเทคโนโลยีประมวลผลภาษาธรรมชาติ (Natural Language Processing) และจัดตามกลุ่มด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means) อยู่ในกลุ่มของการให้คอมพิวเตอร์เรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) ซึ่งเป็นส่วนหนึ่งของเทคโนโลยีปัญญาประดิษฐ์ โดยผลการทดลองใช้พบว่านักศึกษาได้ใช้ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means) (Computer Internship Recommendation System With K-Means Clustering) และได้มีประสิทธิภาพที่ค่าเฉลี่ยค่าความพึงพอใจอยู่ที่ 4.01 ซึ่งอยู่ในระดับดี

## กิจกรรมประจำ

โครงการทางวิทยาการคอมพิวเตอร์ “ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means) (Computer Internship Recommendation System With K-Means Clustering)” เพื่อการสำเร็จการศึกษาระดับปริญญาตรี สามารถดำเนินการจนประสบความสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาและคำแนะนำจากคณาจารย์หลาย ๆ ท่านในหลักสูตรวิทยาการคอมพิวเตอร์ ที่ได้กรุณากล่าวให้ความรู้ แนะนำแนวทาง ข้อคิด ขอแนะนำสู่ความสำเร็จ และช่วยแก้ไขปัญหาอุปสรรคต่าง ๆ รวมทั้งรูปเล่มให้สำเร็จไปได้ด้วยดี

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.นงนุช เกตุย และอาจารย์วรวิทย์ ผึ้นคำข่าย อาจารย์ที่ปรึกษา และอาจารย์ประจำวิชาทุกท่าน ผู้ชี้งกรุณากล่าวให้ความรู้ คำแนะนำแนวทางการสร้างผลงานสู่ความสำเร็จ และช่วยแก้ไขปัญหาอุปสรรคต่าง ๆ รวมทั้งตรวจทานแก้ไขรูปเล่มจนเสร็จสมบูรณ์ ขอขอบคุณบิดา มารดา ผู้มีพระคุณทุกท่าน เพื่อนนักศึกษา และบุคคลที่เกี่ยวข้องที่ยังไม่ได้กล่าวถึง ที่ได้ช่วยออกแบบคิดเห็น ได้ให้ขอแนะนำ และคำนวณความสะดวกในด้านต่าง ๆ ในการทำโครงการครั้งนี้ได้ ณ ที่นี่

สุดท้ายนี้ผู้ศึกษาหวังว่าโครงการฉบับนี้จะเป็นประโยชน์สำหรับมหาวิทยาลัย และนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน และผู้ที่สนใจที่จะศึกษาต่อไป

## สารบัญ

	หน้า
<b>บทคัดย่อ</b>	ก
<b>กิตติกรรมประกาศ</b>	ข
<b>สารบัญ</b>	ค
<b>สารบัญตาราง</b>	ด
<b>สารบัญภาพ</b>	ฉ
<b>บทที่ 1 บทนำ</b>	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
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง</b>	7
2.1 ทฤษฎีและความรู้ที่เกี่ยวข้อง	7
2.2 งานวิจัยที่เกี่ยวข้อง	20
<b>บทที่ 3 ขั้นตอนการดำเนินงาน</b>	23
3.1 การเตรียมและวิเคราะห์ข้อมูล	23
3.2 การทำงานของระบบ	29
3.3 การวิเคราะห์และออกแบบระบบ	29
3.4 การออกแบบฐานข้อมูล	36
3.5 การออกแบบหน้าจอ	38
<b>บทที่ 4 ผลการดำเนินงาน</b>	42
4.1 การวิเคราะห์และการทำ Word segmentation	42
4.2 ขั้นตอนการใช้งานสำหรับผู้ใช้งาน	53
4.3 การวัดค่าความคล้ายคลึง	56
<b>บทที่ 5 สรุปผลการดำเนินงาน</b>	58
5.1 สรุปผลการดำเนินงาน	58
5.2 สรุปปัญหาที่เกิดระหว่างการดำเนินงาน	58
5.3 แนวทางพัฒนาระบบในอนาคต	59
5.4 แบบประเมินความพึงพอใจของผู้ใช้	59
<b>เอกสารอ้างอิง</b>	61
<b>ภาคผนวก ก คู่มือการติดตั้งระบบ</b>	64
1. การจัดกลุ่มข้อมูลด้วยเทคนิคทางปัญญาประดิษฐ์เคมีน (K-Means)	65
2. การสร้างคลัสเตอร์ (Cluster) ของฐานข้อมูลมองโภตีปี (MongoDB) บนเว็บไซต์	68
3. การสร้างฐานข้อมูลในโปรแกรมมองโภตีปี (MongoDB)	70
4. การนำข้อมูลเข้าสู่ฐานข้อมูลมองโภตีปี (MongoDB)	72

	๙
5. การใช้งาน Cloud computing ของ Amazon Web Services	75
6. การติดตั้งและใช้งาน Web API สำหรับคำนวณค่าความคล้ายคลึง (Cosine similarity)	77
7. การติดตั้งและใช้งานเว็บแอปพลิเคชัน (Web application)	78
<b>ภาคผนวก ๖ คู่มือการใช้งาน</b>	<b>81</b>
1. ผู้ดูแลระบบ	82
2. ผู้ใช้งาน	84
<b>ประวัติผู้ศึกษา</b>	<b>56</b>

## 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	23
5. คำอธิบาย Use case คุณมีการใช้งาน	30
6. คำอธิบาย Use case ดูรายชื่อบริษัททั้งหมด	30
7. คำอธิบาย Use case ดูรายชื่อบริษัทในกลุ่มทั้งหมด	31
8. คำอธิบาย Use case ดูข้อมูลบริษัท	31
9. คำอธิบาย Use case คนหาบริษัทด้วยความสนใจ	31
10. คำอธิบาย Use case แก้ไขคุณมีการใช้งาน	32
11. คำอธิบาย Use case เพิ่ม ลบ แก้ไขข้อมูลบริษัท	32
12. อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้	33
13. อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบริษัทใหม่	34
14. พจนานุกรมข้อมูลบริษัท	37
15. แสดงการนับจำนวนบริษัทแต่ละประเภทในการจัดกลุ่ม	49
16. แสดงจำนวน ค่าเฉลี่ยความพึงพอใจต่อระบบ	60
17. ขอเสนอแนะของผู้ใช้	60

## สารบัญภาพ

ภาพที่	หน้า
1. ตัวอย่างการทำ Word segmentation	8
2. การกำหนดสูมกำหนดจุด Centroid	11
3. จุด Centroid ที่อยู่ตรงกลางและจุดข้อมูลทุกจุดไม่เปลี่ยนแปลง	11
4. ตัวอย่างการจัดกลุ่มข้อมูลด้วย K-Means โดยที่เครื่องหมายกากรบทสีแดงคือจุด Centroid ของแต่ละกลุ่มข้อมูล	12
5. ภาพที่แสดงจำนวนผิดพลาดเพื่อหาจพวนกลุ่มที่เหมาะสมที่สุด	12
6. คำนวณค่า Cosine similarity ด้วย scikit-learn	13
7. ตัวอย่างการส่งคำขอไปยัง API	14
8. การตอบกลับจาก API	14
9. ตัวอย่างโค้ดสำหรับการสร้าง Web API ด้วย fastAPI	15
10. ผลลัพธ์แสดงคำว่า Hello project จาก fastAPI	16
11. ตัวอย่างข้อมูลแบบ JSON	17
12. การเตรียมและวิเคราะห์ข้อมูล	24
13. การเรียกใช้ไลบรารี (Library) สำหรับคำนวณค่า TF-IDF	25
14. การอ่านข้อมูลจากไฟล์และกำหนดตัวกรองการตัดคำ	25
15. พังก์ชันสำหรับใช้ลบตัวเลข และอักขระพิเศษ	25
16. พังก์ชันสำหรับใช้ลบคำที่ไม่สื่อความหมายและตัวเลขไทย	26
17. การวนซ้ำข้อมูลเพื่อตัดคำและทำความสะอาดตัวข้อมูล	26
18. การเทรนและการทดสอบโดยเดลการคำนวณค่า TF-IDF	26
19. ผลลัพธ์การคำนวณค่า TF-IDF	27
20. การเรียกใช้ไลบรารีสำหรับการ จัดกลุ่มข้อมูลด้วยเคลมีนท์ (K-Means)	27
21. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล	28
22. การเรียกใช้งานไฟล์ clustering.py เพื่อจัดกลุ่มข้อมูลและบันทึกผลลัพธ์	28
23. การนำเข้าข้อมูลลงสู่ฐานข้อมูล MongoDB	28
24. การทำงานของระบบ	29
25. Use Case Diagram ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลมีนท์ (K-Means)	30
26. Sequence Diagram การคุ้นเคยบริษัทด้วยความสนใจของผู้ใช้	33
27. Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบริษัทใหม่	34
28. Activity Diagram ของผู้ใช้งาน	35
29. Activity Diagram ของผู้ดูแลระบบ	36
30. ER Diagram ระบบแนะนำบริษัทสำหรับฝึกงานตามความสนใจ	37
31. หน้าแรก	38
32. หน้าเกี่ยวกับ	39
33. หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด	39
34. หน้าแสดงรายชื่อบริษัททั้งหมด	40

	๊
35. หน้าแสดงผลลัพธ์รายชื่อบริษัท	40
36. หน้าแสดงข้อมูลบริษัท	41
37. ตัวอย่างผลลัพธ์จากการค้นหาด้วยความสนใจของผู้ใช้	41
38. ตัวอย่างข้อมูลต้นฉบับ	43
39. ตัวอย่างการตัดคำโดยใช้ Engine newmm	43
40. ตัวอย่างการตัดคำโดยใช้ Engine longest	44
41. ตัวอย่างการตัดคำโดยใช้ Engine deepcut	44
42. ผลการวัดค่าความแม่นยำในการตัดคำของ Engine ในไลบรารี Pythainlp	45
43. ตัวอย่างการตัดคำและเรียงคำที่มีน้ำหนักมากที่สุด 5 อันดับ	45
44. ตาราง TF-IDF แสดงน้ำหนักของคำ	46
45. ตัวอย่างการตัดคำและลบ Stop word	46
46. การทำ Elbow method	47
47. จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม	47
48. จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม	48
49. จัดกลุ่มข้อมูลจำนวน 7 กลุ่ม	48
50. จัดกลุ่มข้อมูลจำนวน 6 กลุ่ม	49
51. เปรียบเทียบอัตราการเติบโตการจัดกลุ่มข้อมูล	50
52. ตัวอย่างข้อมูลในกลุ่ม 0	52
53. ตัวอย่างข้อมูลในกลุ่ม 2	52
54. หน้าแรกเว็บไซต์ Intern-assistant	53
55. คนหาบริษัท	53
56. หน้าแสดงผลลัพธ์การค้นหา	54
57. หน้าเกี่ยวกับ	54
58. หน้าแสดงรายชื่อบริษัททั้งหมด	55
59. หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด	55
60. หน้ารายละเอียดบริษัท	56
61. ตัวอย่างการคำนวณค่า Cosine similarity	56
62. ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าความคล้ายคลึง	57
63. ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่คล้ายที่สุด	57
64. ดาวน์โหลด Repository ด้วยคำสั่ง git clone	65
65. แสดงการบันทึกไฟล์ข้อมูลบริษัทในไฟล์เดอร์	65
66. โค้ดคำสั่งในไฟล์ clustering.py ใช้ในการจัดกลุ่มข้อมูล	66
67. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล	66
68. แสดงการรันคำสั่งจัดกลุ่มข้อมูลใน Terminal	67
69. แสดงไฟล์ clustered_company.csv	67
70. แสดงหน้าการจัดการ Cluster MongoDB	68
71. แสดงหน้าตั้งค่าและสร้าง Cluster MongoDB	68
72. แสดงหน้าสร้างบัญชีสำหรับจัดการฐานข้อมูล	69
73. แสดงหน้าเพิ่ม IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้	69

	๊
74. แสดงหน้าจัดการ Cluster MongoDB	70
75. แสดงการสร้างฐานข้อมูล MongoDB	70
76. แสดงหน้าต่างการสร้างฐานข้อมูลและ Collection	71
77. ตัวอย่างการเลือกตั้งค่าการดาวน์โหลดโปรแกรม MongoDB compass	71
78. หน้าต่างเลือกเชื่อมต่อ กับ Cluster	72
79. หน้าต่างข้อมูลการเชื่อมต่อ Cluster กับ MongoDB compass	72
80. หน้าต่างโปรแกรม MongoDB compass สำหรับเชื่อมต่อ Cluster	73
81. หน้าต่างโปรแกรมแสดงข้อมูลใน Collection	73
82. หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv	74
83. หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass	74
84. หน้าเว็บไซต์ Amazon Web Services	75
85. หน้าแสดงการเลือกสร้าง Instance ใหม่	75
86. หน้าแสดงการตั้งค่า Instance	76
87. ตัวอย่างการเชื่อมต่อเข้าไปยัง Instance	76
88. การดาวน์โหลดโปรเจคจาก Github ด้วยคำสั่ง git clone มา�ัง Instance และใช้คำสั่ง cd final_project เพื่อเข้าไปยังโฟลเดอร์	77
89. สร้างไฟล์ใหม่ชื่อ .env และสร้างตัวแปรเพื่อเก็บค่าเชื่อมต่อฐานข้อมูล	77
90. ตัวอย่างการรัน Server Web API เพื่อคำนวณค่า Cosine similarity บน AWS	78
91. หน้าเว็บไซต์ Vercel	78
92. สร้างโปรเจคใหม่ใน Vercel	79
93. หน้าแสดงรายชื่อ Repository	79
94. หน้าการตั้งค่าโปรเจคก่อน Deploy	80
95. ตัวอย่างหน้าเว็บไซต์	80
96. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k	82
97. แสดงการใช้คำสั่งจัดกลุ่มข้อมูลใน Terminal	82
98. แสดงไฟล์ clustered_company.csv	83
99. หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv	83
100. หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass	84
101. หน้าเว็บไซต์ intern-assistant.vercel.app	84
102. หน้าแสดงผลลัพธ์เมื่อคนหาบริษัท	85
103. หน้าแสดงข้อมูลบริษัท	85
104. วิธีการส่งคำขอไปยัง https://iamonze.tech/allcompanies	86
105. วิธีการส่งคำขอไปยัง https://iamonze.tech/cluster/1	86

	๙
106.	๘
วิธีการส่งคำขอไปยัง <a href="https://iamonze.tech/cluster">https://iamonze.tech/cluster</a>	๘๗
107.	๙
วิธีการส่งคำขอไปยัง <a href="https://iamonze.tech/search">https://iamonze.tech/search</a>	๘๗
108.	๙
วิธีการส่งคำขอไปยัง <a href="https://iamonze.tech/searchcompany">https://iamonze.tech/searchcompany</a>	๘๘

## บทที่ 1 บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในระบบการศึกษาระดับปริญญาตรีนั้นรายวิชาที่มีในการศึกษาปีสุดท้ายของหลักสูตรคือ รายวิชาที่จะต้องให้นักศึกษาแต่ละคนนั้นออกแบบ自己ทำงานที่สถานประกอบการต่างๆ ในช่วงระยะเวลาหนึ่ง ซึ่งเป็นสิ่งที่มีความสำคัญและความท้าทายเนื่องจากเป็นการที่นักศึกษาจะได้ทดลองทำงานจริง สถานการณ์จริง สถานที่จริง ในสถานประกอบการที่นักศึกษาได้เลือก

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

ด้วยเหตุนี้จึงได้มีการเริ่มโครงการพัฒนาเว็บแอปพลิเคชันที่นักศึกษามีความสนใจในรูปแบบธุรกิจของสถานประกอบการนั้น เพื่อช่วยอำนวยความสะดวกแก่นักศึกษา มหาวิทยาลัย เทคโนโลยีราชมงคลล้านนา น่าสนใจ การพัฒนาสถานประกอบการสำหรับออกฝึกงาน 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 เพื่อจัดการคำและแบ่งกลุ่มข้อมูลโดยใช้ เทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means) และหาความคล้ายของข้อมูลด้วยการคำนวณความ คล้ายคลึง (Cosine similarity)

## 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

การศึกษาค้นคว้าเพื่อจัดทำโครงการทางวิทยาการคอมพิวเตอร์ ระบบแนะนำสถานที่ฝึกงาน ด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อนที่ (K-Means) ผู้ศึกษาได้ ศึกษาค้นคว้าเอกสาร และงานวิจัยที่เกี่ยวข้องตามหัวข้อลำดับต่อไปนี้

#### 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 เป็นเทคโนโลยีในเทคโนโลยีปัญญาประดิษฐ์ ที่จะทำให้คอมพิวเตอร์สามารถเข้าใจและเรียนรู้ ประมวลผลภาษาของมนุษย์ได้ ในด้านของการวิเคราะห์ภาษาศาสตร์ การตีความจากบทความ หรือกระทั่งการแปลภาษา 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 นั้นมีเครื่องมือให้เช่นอย่างจำนวนมากยกตัวอย่าง เช่น Python library pythainlp, nltk หรือสามารถใช้บริการ web API ของ aiforthai

```
onze@Tinngrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh)  #1 ..roject/report (-zsh)  #2 +-
25% 4.3 GB main + * ~Desktop/final_project
└ python report.py
['น', 'อน', 'ต', 'า', 'ก', 'ล', 'ม', 'ด', 'ู', 'دا', 'ว']
```

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

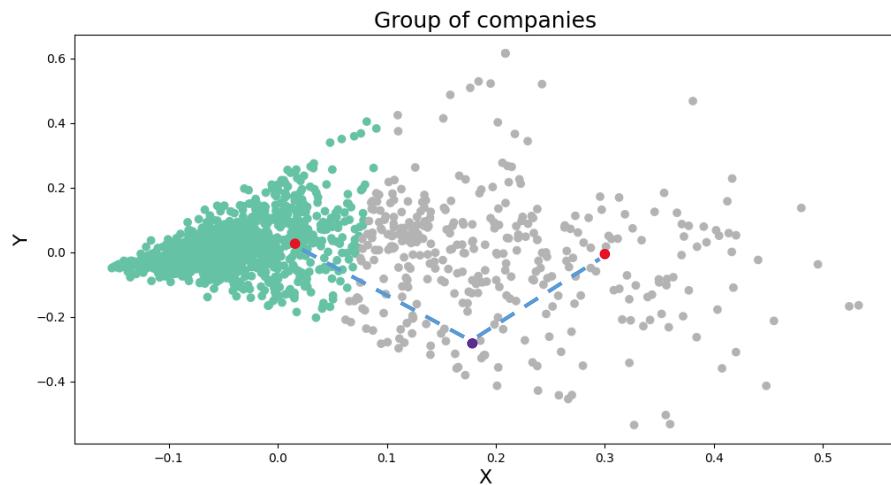
ดังตัวอย่างจะเห็นได้ว่าบางคำที่มีค่า TF-IDF สูงแต่ไม่ได้บ่งบอกถึงลักษณะของข้อความในเอกสาร เช่นคำว่า และ ซึ่งถือว่าเป็น Stop word ซึ่งเป็นคำที่ไม่สื่อความหมายโดยปกติแล้ว คำเหล่านี้มักถูกกรองออกก่อนที่จะมีการนำข้อมูลมาทำการประมวลผลทางภาษา自然ภาษาตัวอย่างจะเห็นว่าเมื่อคำนวณหาค่า Term Frequency – Inverse Document 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)

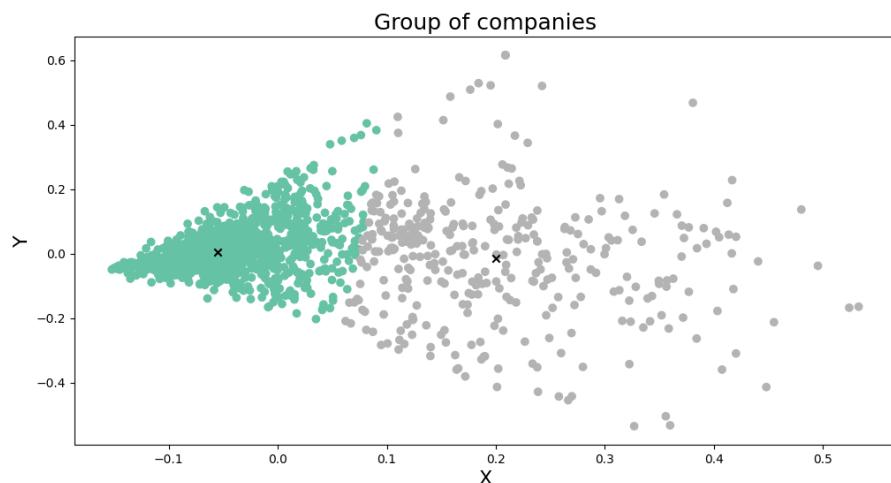
$$Distance = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

ตัวอย่างการคำนวณค่าการกำหนดจุด Centroid เมื่อกำหนด k=2 ทำการสูมจุดข้อมูลจากข้อมูลทั้งหมด



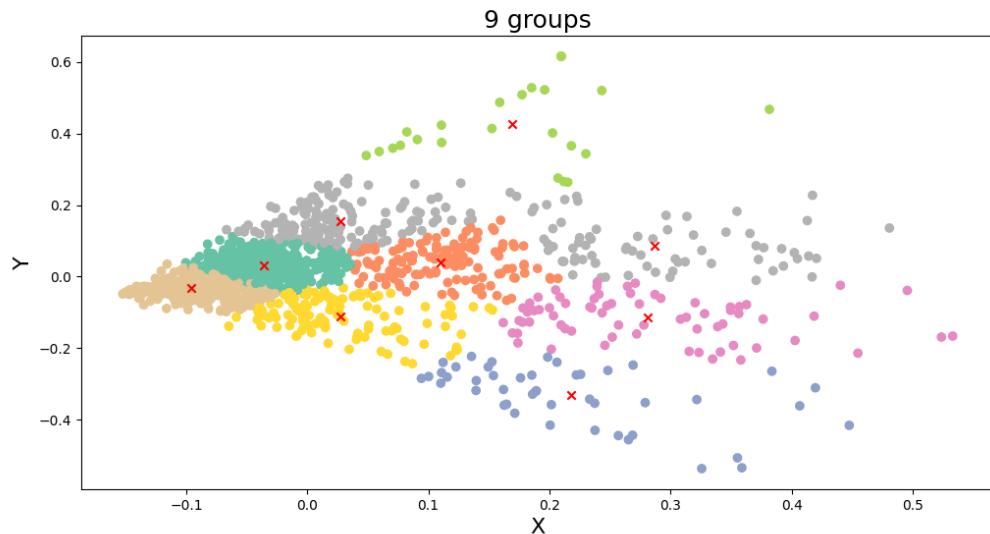
ภาพที่ 2 การกำหนดสูมกำหนดจุด Centroid

ทำการทำซ้ำการกำหนดจุด centroid จนกว่าตำแหน่งของข้อมูลทุกตัวจะไม่เปลี่ยนแปลง



ภาพที่ 3 จุด Centroid ที่อยู่ตรงกลางและจุดข้อมูลทุกจุดไม่เปลี่ยนแปลง

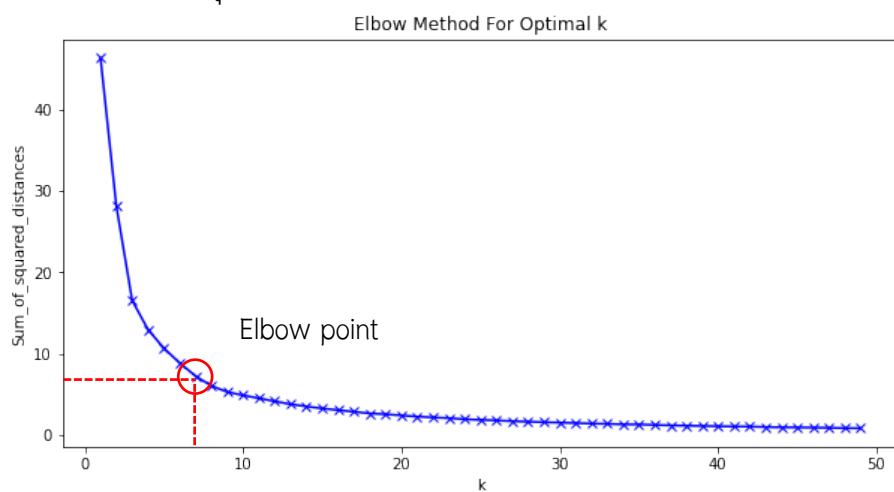
ทำการคำนวณและขยับจุด Centroid และหาค่าเฉลี่ยจนค่าเฉลี่ยไม่มีการเปลี่ยนแปลงจะได้จุดกึ่งกลางของข้อมูลในแต่ละกลุ่ม



ภาพที่ 4 ตัวอย่างการจัดกลุ่มข้อมูลด้วย K-Means โดยที่เครื่องหมายกากรากที่สีแดงคือจุด Centroid ของแต่ละกลุ่มข้อมูล

#### 2.1.5 ทฤษฎี การหาจำนวนกลุ่มที่เหมาะสมสมด้วยวิธี Elbow method

Elbow method เป็นวิธีหนึ่งที่ใช้หาจำนวนของกลุ่มที่เหมาะสมสมด้วยการวัดข้อผิดพลาด (Error measurement) ผลรวมระยะห่างระหว่างข้อมูลกับจุด Centroid เมื่อข้อผิดพลาดน้อยลงความชันของเส้นคงที่จะแบบราบไปตามแกน X จนทำให้เกิดมุมลักขณา晦เมื่อนอกจากจะถือว่าจุดที่อยู่ตรงมุมของคอกเป็นจำนวนของกลุ่มข้อมูลที่เหมาะสมสมดังในภาพต่อไปภาพที่ 3 จะเห็นได้ว่าจำนวนกลุ่มที่เหมาะสมสมคือ 4-5 กลุ่ม (Paul, 2021)



ภาพที่ 5 กราฟที่แสดงจำนวนข้อผิดพลาดเพื่อหาจำนวนกลุ่มที่เหมาะสมสมที่สุด

#### 2.1.6 ทฤษฎี การคำนวณค่าความคล้ายคลึงด้วยเทคนิค Cosine similarity

การวัดความเหมือนของ Vector 2 (Cosine Similarity) ว่าปြုไปในทิศทางเดียวกันหรือไม่โดยที่เป็นการตัดขนาด หรือ Magnitude ของ Vector ออกไปหากค่าได้จากสมการนี้

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

อธิบายโดยง่ายคือเป็นการวัดระยะห่างระหว่าง Object A และ Object B ว่ามีความคล้ายกันแค่ไหนยกตัวอย่างเบรียบเทียบระหว่างคำว่า “ยินดีที่ได้รู้จักรับ” และ “ยินดีที่ได้รู้จัคคง” ทำการตัดเพื่อหาคำทั้งหมดก่อนครับ คือ [ “ยินดี” , “ที่” , “ได้” , “รู้จัก” , “ครับ” , “คง” ] เรายจะได้ Object A และ B เป็นชุดข้อมูล ได้ดังนี้ (Supalerk, 2020)

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

```
onze@Tinngrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh)          #1           ..project/report (-zsh)      #2 +
27% 4.7 GB | main + * | ~/Desktop/final_project
└ python report.py
Cosine similarity: [0.8]
~/desktop/final_project main !1 ?3 → ✓ backup_finalproject 14:09:00
```

ภาพที่ 6 คำนวณค่า Cosine similarity ด้วย scikit-learn

$$\text{similarity} = \frac{4}{\sqrt{(1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2)} \times \sqrt{(1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 1^2)}}$$

$$\text{similarity} = \frac{4}{0.8} = 0.8$$

### 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 เว็บไซต์ซื้อขายลินค้าออนไลน์ชื่อดังจากประเทศไทย 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 คือกลไกที่ช่วยให้ส่วนประกอบ

ซอฟต์แวร์สองส่วนสามารถสื่อสารกันได้โดยใช้ชุดคำจำกัดความและโปรโตคอล ตัวอย่างเช่น ระบบซอฟต์แวร์ของสำนักพยากรณ์อากาศประกอบด้วยข้อมูลสภาพอากาศรายวัน (API คืออะไร)

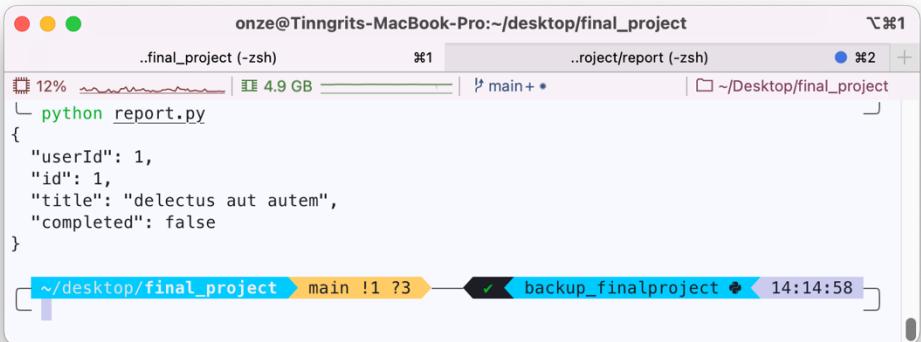


```

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

```

ภาพที่ 7 ตัวอย่างการส่งคำขอไปยัง API



```

onze@Tinngrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh)   *1   ..project/report (-zsh)   *2   + 
└ python report.py
{
  "userId": 1,
  "id": 1,
  "title": "delectus aut autem",
  "completed": false
}

```

ภาพที่ 8 การตอบกลับจาก API

### 2.1.9 ทุกเรื่อง Cloudflare

คลาวด์เฟร์ (Cloudflare) คือ Global Network ที่ถูกออกแบบมาเพื่อให้ทุกสิ่งของคุณที่เชื่อมอยู่บนอินเทอร์เน็ต 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 จุดต่างๆ หากมี Traffic ระบบก็จะส่งข้อมูลโดยใช้ Server ที่อยู่ใกล้ที่สุด โดย Cloudflare มี POPs ในไทยมากถึง 6 POPs และมากกว่า 200 POPs ทั่วโลก ช่วยให้เว็บไซต์สามารถใช้งานได้อย่างรวดเร็ว และเสียรุ่น

#### 2.1.10 ทฤษฎี Cors

การอนุญาตการแบ่งปันข้อมูลกัน (Cross-Origin Resource Sharing:CORS) เป็นกลไกที่ใช้เพิ่มเติมเพื่อให้บรรดาเซอร์ฟเวอร์ได้รับสิทธิในการเข้าถึงทรัพยากรที่เลือกจากเซิร์ฟเวอร์บนโดเมนอื่นมาแสดงบนหน้าเว็บของบรรดาเซอร์ฟเวอร์ได้ คอมพิวเตอร์แต่ละเครื่องทั้งที่มี Protocol ที่เหมือนกัน ถึงจะลืมรหัสผ่านไว้ เว็บบรรดาเซอร์ฟเวอร์จะส่ง HTTP request เมื่อต้องการขอข้อมูลข้ามโดเมนหรือ port ที่ต่างกัน และต้องทำตามข้อตกลงการสื่อสาร (Protocol) เพราะปัจจุบันเรามักจะแยกฝั่ง 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 และมีเจ้าตัว Uvicorn เป็นตัว run server ของดีของการใช้งาน fastAPI คือ (Natakorn, 2021)

1. มีความเร็วของการทำงานเทียบเท่า Node.js และ Go
2. รูปแบบการเขียนฟังก์ชันต่างๆ เช่นง่ายต่อการศึกษา
3. ง่ายต่อการใช้งานและพัฒนาต่อ

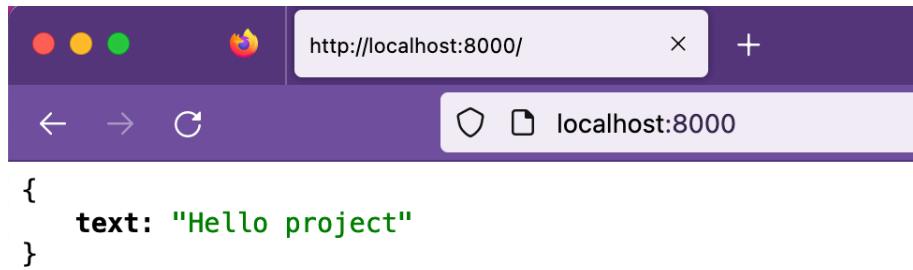


```

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

```

ภาพที่ 9 ตัวอย่างโค้ดสำหรับการสร้าง Web API ด้วย fastAPI



```
{
    text: "Hello project"
}
```

ภาพที่ 10 ผลลัพธ์แสดงคำว่า Hello project จาก fastAPI

### 2.1.12 ทฤษฎี Git

Git คือ Version Control ที่ถูกพัฒนาขึ้นมาเพื่อใช้ในกระบวนการพัฒนาซอฟต์แวร์ แปลให้เข้าใจแบบง่าย ๆ คือ ระบบที่ถูกพัฒนาขึ้นมาเพื่อใช้สำหรับการติดตาม ตรวจสอบ การพัฒนา แก้ไข ซอฟต์แวร์ ซึ่งมีความสามารถในการตรวจสอบได้ทุกตัวอักษร ทุกบรรทัด ทุกไฟล์ ที่มีการแก้ไข โครงเป็นคนแก้ไข และแก้ไข ณ วันที่เท่าไหร่

ระบบการทำงานของ Git ไม่ได้อยู่แค่การตรวจสอบการแก้ไขเท่านั้น ยังสามารถรวมการแก้ไขทั้งหมดเข้าด้วยกันได้อย่างชั้นตอนนี้ ที่เรียกว่า CI ( Continuous Integration ) และในปัจจุบัน Git VCS ( Version Control System ) มีการควบรวมไฟล์เข้าด้วยกัน ทำให้นักพัฒนาทำงานได้สะดวกมากขึ้น สามารถทำงานได้ตั้งแต่ชั้นตอนการพัฒนา ไปจนถึงการ Deploy งานขึ้นใช้งานบน Server เราเรียกว่า CD ( Continuous Deployment ) รูปแบบการใช้งานของ Git มีด้วยกัน 2 รูปแบบ คือ (codebee, 2020)

- ใช้งานผ่าน Git Command Line ( ใช้งานผ่านการพิมพ์คำสั่งด้วยหนังสือ )
- ใช้งานผ่านโปรแกรม Git GUI ( ใช้งานผ่านโปรแกรมสำเร็จรูป )

### 2.1.13 ทฤษฎี Node.js

Node.js คือสภาพแวดล้อมการทำงานของภาษา JavaScript นอกเว็บเบราว์เซอร์ที่ทำงานด้วย V8 engine นั้นหมายความว่าเราสามารถใช้ Node.js ในการพัฒนาแอ��พพลิเคชันแบบ Command line และพลิกเคชัน Desktop หรือแม้แต่เว็บเฟอร์ได้ โดยที่ Node.js จะมี APIs ที่เราสามารถใช้สำหรับทำงานกับระบบปฏิบัติการ เช่น การรับค่าและการแสดงผล การอ่านเขียนไฟล์ และการทำงานกับเน็ตเวิร์ก เป็นต้น

Node.js ถูกพัฒนาและทำงานด้วย Chrome V8 engine สำหรับคอมโพล์ภาษา JavaScript ให้เป็นภาษาเครื่องด้วยการคอมไพล์แบบ Just-in-time (JIT) เพื่อเพิ่มประสิทธิภาพการทำงานของภาษา JavaScript จากที่แต่เดิมมันเป็นภาษาที่มีการทำงานแบบ Interpreted Node.js เป็นโปรแกรมที่สามารถใช้ได้ทั้งบน Windows, Linux และ Mac OS X นั้นหมายความว่าคุณสามารถเขียนโปรแกรมในภาษา JavaScript และนำไปรันได้ทุกระบบปฏิบัติการที่สนับสนุนโดย Node.js นี่เป็นแนวคิดของการเขียนครั้งเดียวแต่ทำงานได้ทุกที่ (Write once, run anywhere) ขอต้องอย่างหนึ่งในการใช้ภาษา JavaScript ของ Node.js คือทำให้การพัฒนาเว็บไซต์ทำได้อย่างชั้นสำหรับนักพัฒนา เนื่องจากความสามารถใช้ภาษา JavaScript สำหรับทั้ง Front-end และ Back-end ได้โดยไม่ต้อง

ศึกษาภาษาเฉพาะในแต่ละด้าน ตัวอย่างของการพัฒนาเว็บไซต์ในรูปแบบนี้ เช่น React.js ซึ่งเป็นไลบรารีโดย Facebook (ทำความรู้จักกับ Node.js, 2021)

#### 2.1.14 ทฤษฎี Matplotlibs

Matplotlib เป็นโมดูลที่เป็นพื้นฐานของ Python สำหรับการวิเคราะห์ข้อมูลซึ่งจำเป็นมากสำหรับงานทางด้าน Data Analysis, Science, Engineering เป็นตัวช่วยในการวิเคราะห์ข้อมูลโดยใช้รูปแบบของกราฟตัวอย่างประเภทกราฟที่มีให้ใช้ 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 คือไม่มี relation (ความสัมพันธ์) ของตารางแบบ SQL ทั่วไป แต่จะเก็บข้อมูลเป็นแบบ JSON (JavaScript Object Notation) แทน การบันทึกข้อมูลทุกๆ record ใน MongoDB เราจะเรียกมันว่า Document ซึ่งจะเก็บค่าเป็น key และ value จะเห็นว่ามันก็คือ JSON นั่นแหล่ะ (Chai, 2015) ตัวอย่างเช่น



```

1 {
2   "userId": 1,
3   "id": 1,
4   "title": "delectus aut autem",
5   "completed": false
6 },
7 {
8   "userId": 1,
9   "id": 2,
10  "title": "quis ut nam facilis et officia qui",
11  "completed": false
12 },

```

ภาพที่ 11 ตัวอย่างข้อมูลแบบ JSON

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

1. ตัว MongoDB สามารถที่จะสร้างเป็น Cluster เพื่อที่จะตอบสนองของคำว่า High Availability (HA) ได้ ซึ่งเราอาจจะเลือก Region ที่เรอယากจะ Deploy บน Cloud Provider นั้นๆ ได้
2. ความรวดเร็วในการเข้าถึงข้อมูล เพราะว่าตัว Database ของมันไม่มี Schema ซึ่งจะต่างกับ SQL โดยพกนั้นจะอิงจากฐานข้อมูลที่มาจากการ Table
3. สามารถทำ Auto Scale ได้ไม่ว่าจะมีการใช้งานมากน้อยแค่ไหน ตัวมันก็สามารถ 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 ให้เรียบง่ายครบถ้วน 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 เป็นที่นิยมในหลายๆ บริษัท นอกจากนี้ ข้อดีก็คือ render ได้เร็วกว่า React เพราะ Next.js มีสิ่งที่เรียกว่า get static path ซึ่งการสร้าง path แบบ static แบบเว็บไซต์ html โดยไม่ต้องทำการเชื่อมต่อกับ back end เพื่อให้ได้ data ยังไงก็ได้ สำหรับ Next.js สามารถรวมเข้ากับ backend ได้ง่ายๆ เพราะ Next.js มีสิ่งที่เรียกว่า API routes ใน การรับส่ง request ใน folder ของ page จะมีอีก folder ที่เรียกว่า API ที่ถูกปฏิบัติเป็น endpoint แทนที่จะเป็น page ซึ่ง folder API นี้จะเป็นในส่วนหนึ่งของ server-side เท่านั้น ทำให้ไม่โปะเพิ่ม size ของ client side (frevation, 2021)

### 2.1.17 ทฤษฎี Numpy

NumPy เป็น Library(ไลบรารี) ที่รุ่นจัก และเป็นที่นิยมใช้ในการคำนวณเช่น ใช้คำนวณ Matrix หรือ คำนวณกับ Array ในงาน Data Science, Data analytics และในการทำ Machine Learning (การเรียนรู้ของเครื่องจักร) หรือ Deep Learning (ดีป เลิฟนิ่ง) ก็ยังคงต้องใช้ NumPy ออยู่ดี NumPy เป็น Library พื้นฐานที่ใช้คำนวณทางคณิตศาสตร์ด้วยภาษา Python สามารถคำนวณ หรือดำเนินการทำตระกรากใน 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) เมตริกซ์ (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 และวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพตั้งแต่ข้อมูล

ขนาดเล็กไปจนถึงข้อมูลขนาดใหญ่ ทำให้ pandas ตอบโจทย์งานในยุคที่ข้อมูลมีขนาดใหญ่มากขึ้น เอื้อย ๆ ได้ ไม่มีปัญหาติดขัดเมื่อกับ Spreadsheets อื่นๆ (เช่น Excel หรือ Google Sheets ซึ่งจะทำงานได้ช้าลงหากข้อมูลมีขนาดใหญ่ขึ้น) ขั้นตอนการเตรียมข้อมูลนั้นมีความสำคัญมาก และ Data Scientist อาจจะใช้เวลาส่วนใหญ่หมดไปกับขั้นตอนนี้ เพราะหากข้อมูลที่เตรียมได้ไม่มีประสิทธิภาพ การนำ Insights ไปใช้งาน หรือนำข้อมูลไปสร้างโมเดล ย่อมทำให้ได้ข้อมูลที่ไม่ง่ายเชื่อถือนั่นเอง

นอกจากนี้ เมื่อเปรียบเทียบกับ Tools วิเคราะห์ข้อมูลอื่นๆ อย่าง Excel หรือ Google Sheets อาจไม่ตอบโจทย์เต็มที่หากต้องการเชื่อมต่อกับแหล่งข้อมูลบางประเภท หรือทำ Automation (ระบบจัดการอัตโนมัติ) ในขณะที่ pandas ซึ่งเป็นส่วนหนึ่งของ Python นั้นสามารถใช้การเขียนโค้ด เพื่อปรับแต่ง 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 มีการแสดง เกี่ยวกับเซ็ตของภาษาไทยทั้งหมด ยังมีในส่วนของการแยกตัวอักษร เป็นส่วนของพยัญชนะ สรุว วรรณยุกต์ เป็นต้น ยังมีในส่วนของเลขไทย มีการเช็คว่าเป็นคำภาษาไทยรึไม่ มีการวนับตัวอักษร ว่าเป็นภาษาไทยกี่เปอร์เซ็นต์ และยังมีส่วนของการแสดงคำอ่านที่เป็นพากเวลาและยังมีการจัดเรียงคำใน List ให้เรียงกันเป็นลำดับได้ เป็นต้น (mindphp, 2022) นอกจากนี้ยังใช้สำหรับประมวลผลข้อความ 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 สำหรับการดำเนินการพิเศษเชิงเส้นและอาร์เรย์ที่มีประสิทธิภาพสูง (เจร์, 2021)

Scikit-Learn ถูกสร้างขึ้นโดยเป็นส่วนหนึ่งของโครงการ Summer of Code ของ Google และทำให้ชีวิตของนักวิทยาศาสตร์ข้อมูลที่มี Python เป็นศูนย์กลางนับล้านทั่วโลกง่ายขึ้น ส่วนนีของซีรีส์มุ่งเน้นไปที่การนำเสนอไลบรารีและมุ่งเน้นไปที่องค์ประกอบเดียว นั่นคือการแปลงชุดข้อมูล ซึ่งเป็นขั้นตอนสำคัญและสำคัญที่ต้องทำก่อนพัฒนาแบบจำลองการทำงาน 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 โดยดูของคุณได้อย่างง่าย Vercel Inc. เติมชื่อ Zeit เป็นแพลตฟอร์มคลาวด์ของอเมริกาในฐานะบริษัทผู้ให้บริการ บริษัทรักษากรอบการพัฒนาเว็บไซต์ Next.js สถาปัตยกรรมของ Vercel สร้างขึ้นจาก Jamstack และการจัดการการปรับใช้ผ่านที่เก็บ Git Vercel เป็นสมาชิกของ MACH Alliance (Huangsri, 2021)

## 2.2 งานวิจัยที่เกี่ยวข้อง

จักรินทร์ สันติรัตนภักดี และศุภกฤษฐ์ นิวัฒนาภูล ศึกษาเรื่อง การออกแบบและพัฒนากระบวนการจำแนกข้อร้องเรียนรถโดยสารสาธารณะเพื่อติดแท็กปัญหาการให้บริการ องค์การขนส่งมวลชนกรุงเทพ (ขสมก.) มีช่องทางในการร้องเรียนรถโดยสารสาธารณะผ่านเว็บบอร์ด ที่ผู้ใช้งานสามารถแสดงความคิดเห็นได้อย่างอิสระ ผู้วิจัยจึงออกแบบและพัฒนากระบวนการจำแนกข้อร้องเรียนรถโดยสารสาธารณะ จากข้อร้องเรียนผ่านเว็บบอร์ดขององค์การขนส่งมวลชนกรุงเทพด้วยกระบวนการตัดคำภาษาไทยโดยใช้พจนานุกรม แล้วคัดเลือกคำพท์ด้วยการวิเคราะห์น้ำหนักของคำมาสร้างเป็นคลังคำพท์ แบ่งเป็น 4 คลาส ได้แก่ คลาสการขับขี่ คลาสผู้ขับขี่และพนักงานผู้ให้บริการ คลาสขานพาณะและอุปกรณ์ให้บริการ และคลาสวุฒิภาวะเดินรถโดยใช้ชุดลักษณะตัดคำภาษาไทย (Thai Word Segmentation) ด้วยข้อความทั่วไปซึ่งอยู่ในรูปแบบประโยคมาแบ่งออกเป็นคำหรือคุณลักษณะ (Term/Feature) เพื่อแยกส่วนของข้อความออกจากกันก่อนนำไปประมวลผลในขั้นต่อไป แบ่งตามกระบวนการทำงานออกเป็น 3 กลุ่ม ได้แก่ 1) การตัดคำโดยใช้กฎ (Rule-Based Approach) 2) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) 3) การตัดคำโดยใช้คลังคำพท์ (Corpus-Based Approach) จากการทดลองพบว่าอัลกอริทึมโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวมสูงที่สุด (จักรินทร์ สันติรัตนภักดี, 2021)

วุฒิชัย วิเชียรไชย ศึกษาเรื่อง การประยุกต์ใช้บริการแบ่งแยกคำภาษาไทยด้วยโครงสร้างการเขียนกับโครงสร้างพยานค์ งานวิจัยนี้นำเสนอการแบ่งแยกคำภาษาไทยโดยเทียบกับโครงสร้างการเขียนของภาษาไทยและอัลกอริทึมการแบ่งแยกคำภาษาไทยโดยโครงสร้างพยานค์ เพื่อศึกษาและประยุกต์ใช้บริการประมวลผลของการแบ่งแยกคำภาษาไทยและประสิทธิภาพความถูกต้องของอัลกอริทึม โดยสามารถแบ่งงานวิจัยในการแบ่งแยกคำภาษาไทยโดยโครงสร้างพยานค์ เป็นคู่ศึกษาและเปรียบเทียบวิธีการประมวลผลของการแบ่งแยกคำภาษาไทยและประสิทธิภาพความถูกต้องของอัลกอริทึม โดยสามารถแบ่งงานวิจัยในการแบ่งแยกคำภาษาไทยโดยโครงสร้างพยานค์ เป็นคู่ศึกษาและเปรียบเทียบวิธีการประมวลผลของการแบ่งแยกคำภาษาไทยและประสิทธิภาพความถูกต้องของอัลกอริทึม โดยโครงสร้างการเขียนภาษาไทยเพื่อแก้ไขลดพื้นที่ในการจัดเก็บคำพท์ในพจนานุกรม และวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างพยานค์เพื่อลดการสิ้นเปลืองพื้นที่ในการจัดเก็บ

พจนานุกรม ยกตัวอย่างการแบ่งแยกคำและพยางค์ของคำว่า “ประเทศไทย” จะสามารถแบ่งแยกคำได้เป็น “ประเทศไทย” และแบ่งพยางค์ได้เป็น “ประเทศไทย” จากผลลัพธ์ในการแบ่งแยกคำนั้นยังขาดความถูกต้องในการแบ่งแยกคำซึ่งสามารถพัฒนาแนวคิดในการศึกษาและสร้างกฎเพื่อแบ่งแยกคำให้ถูกต้องมากยิ่งขึ้น (วุฒิชัย, 2013)

ปราณีํ พึงวิชา อานันท์ ทับเที่ยง และชัญญา สัตยาภิธาน (2019) ศึกษาการแบ่งกลุ่มพฤติกรรมของผู้บริโภคที่ชื่อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ ผู้วิจัยเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจำนวน 400 คน ทำการวิเคราะห์แบ่งกลุ่ม ผู้บริโภคด้วยวิธี K-Means Clustering เป็น 2 กลุ่มซึ่งมีลักษณะเฉพาะในแต่ละกลุ่ม จากการวิเคราะห์ความแตกต่างระหว่างกลุ่มของหัตถศิลป์ด้านพฤติกรรมการซื้อและด้านส่วนประสมทางการตลาดที่มีผลต่อการตัดสินใจซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์เมื่อยุคต่างกัน โดยการวิเคราะห์ความแปรปรวนทางเดียว(One-way ANOVA) พบว่า ด้านพฤติกรรมการซื้อทั้ง 2 กลุ่ม มีความถี่ในการซื้อต่างๆ แฟชั่นไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนตัวแปรอื่น ๆ นั้นมีความแตกต่างกัน จากการวิเคราะห์แบ่งกลุ่มผู้บริโภคที่ชื่อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ด้วยวิธี K-mean clustering สามารถจำแนกเป็น 2 กลุ่มโดยแต่ละกลุ่มมีลักษณะเฉพาะดังนี้ กลุ่มที่1 : กลุ่มกระเปาหนักจ่ายได้ถ้าชอบ ไม่ค่อยชอบออกเสื้อ ลักษณะด้านประชากรศาสตร์ โดยส่วนใหญ่เป็นคนGeneration X เพศหญิงมากกว่าเพศชายมีระดับการศึกษาสูงกว่าปริญญาตรี กลุ่มที่2 : กลุ่มวัยสะอ้อน ชอบออกเสื้อ ชื่อน้อยแต่บ่อยครั้ง ลักษณะด้านประชากรศาสตร์ โดยส่วนใหญ่เป็นคนGeneration Y เพศหญิงมากกว่าเพศชาย ส่วนใหญ่มีระดับการศึกษาสูงแต่จะน้อยกว่ากลุ่ม 1 โดยมีระดับปริญญาตรีมากที่สุดโดยส่วนใหญ่เป็นพนักงานบริษัทเอกชน (ปราณีํ พึงวิชา, 2019)

ธงชัย คล้ายคลึง วุฒิชัย สง่างาม กิตติวงศ์ สุธรรมโน และพันธ์พงศ์อภิชาตกุลศ (2019) ศึกษาเรื่อง เทคนิคการคัดเลือกกลุ่มโหลดรายอาหารสำหรับรับแพนการติดตั้งระบบผลิตไฟฟ้า พลังงานแสงอาทิตย์บนหลังคาเพื่อเพิ่มค่าธรรมเนียมประสิทธิภาพการใช้พลังงานไฟฟ้า บทความนี้ต้องการนำเสนอเทคนิคหรือการคัดเลือกกลุ่มโหลดในแต่ละอาคารที่มีความเหมาะสมสำหรับติดตั้งระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์กรณีที่ติดตั้งบนหลังคาของอาคารในศูนย์กลางมหาวิทยาลัยเทคโนโลยีราชมงคลชีวานนคราชสีมา ในวิธีการของ K-Means Clustering เริ่มนั่นด้วยการจัดแบ่งข้อมูลออกเป็น K กลุ่ม กำหนดจุดศูนย์กลางเริ่มนั่นจำนวน K จุดขึ้นตอนต่อไปคือการสร้างกลุ่มข้อมูลและความสัมพันธ์กับจุดศูนย์กลางที่ใกล้มากที่สุด จากผลการวิเคราะห์การใช้พลังงานไฟฟ้าทั้งหมด 34 อาคารด้วยวิธีการ K-Mean Clustering ทำให้สามารถแยกแยะจัดกลุ่มโหลดรายอาหารได้อย่างมีประสิทธิภาพซึ่งแบ่งได้ 3 กลุ่มโดยกลุ่มที่ 3 จำนวน 19 อาคารนั้นเป็นกลุ่มอาคารที่มีความเหมาะสมทั้งด้านพฤติกรรมการใช้พลังงานไฟฟ้าและมีพื้นที่รองรับการติดตั้งระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์บนหลังคาได้ (Thongchai Klayklueng, 2019)

วรารการ ประดิษฐ์กุล ปราลี มณีรัตน์ และ นิเวศ จิระวิชิตชัย (2021) ศึกษาเรื่อง ระบบแนะนำรายนต์ให้กับลูกค้าโดยการวิเคราะห์จากการซื้อของถึงพฤติกรรมของผู้ใช้ (Collaborative Filtering) กรณีศึกษาบริษัท โตโยต้า บัสส์ จำกัด ผู้วิจัยได้พัฒนาระบบแนะนำรายนต์ให้กับลูกค้าโดยการวิเคราะห์จากการซื้อของถึงพฤติกรรมของผู้ใช้ เพื่อช่วยให้ลูกค้าได้รับการแนะนำรุ่นรถยนต์ที่เหมาะสม ตรงตามความต้องการของลูกค้า ผู้วิจัยใช้อัลกอริทึมการหาความคล้ายคลึงกันของผู้ใช้โดยวิเคราะห์จากลูกค้าที่มีพฤติกรรมใกล้เคียงกันด้วยสมการการหาความคล้ายคลึงโคไซน์ ซึ่งเป็นฟังก์ชันในภาษา Python ในการพัฒนาระบบแนะนำรายนต์ให้กับลูกค้า ด้วยสมการความคล้ายโคไซน์ (cosine similarity) จากการทดลองเมื่อนำข้อมูลมาจัดลำดับคะแนนความชอบของผู้ใช้แต่ละคน เพื่อเป็นการเพิ่มความเร็วให้อัลกอริทึมของวิธีการกรองแบบร่วมมือ อีกทั้งระบบจะนำความคล้ายคลึง

โคโซน์ของผู้ใช้ในระบบกับผู้ใช้เป้าหมายมาทดสอบความแม่นยำของระบบด้วยค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ พบร่วมกัน 0.97 เมื่อกำหนดค่า  $k$  ไว้เท่ากับ 5 สรุปได้ว่าระบบมีประสิทธิภาพในการแนะนำร้านค้าที่รวดเร็วและความแม่นยำอยู่ในระดับที่ดี (Warakorn Pradiskul, 2021)

### บทที่ 3 ขั้นตอนการดำเนินงาน

การจัดทำโครงการทางวิทยาการคอมพิวเตอร์ ระบบแนะนำบริษัทสำหรับผู้ประกอบการตามความสนใจด้วยเทคโนโลยีปัญญาประดิษฐ์ ผู้จัดทำจึงได้กำหนดขั้นตอนการดำเนินงานดังนี้

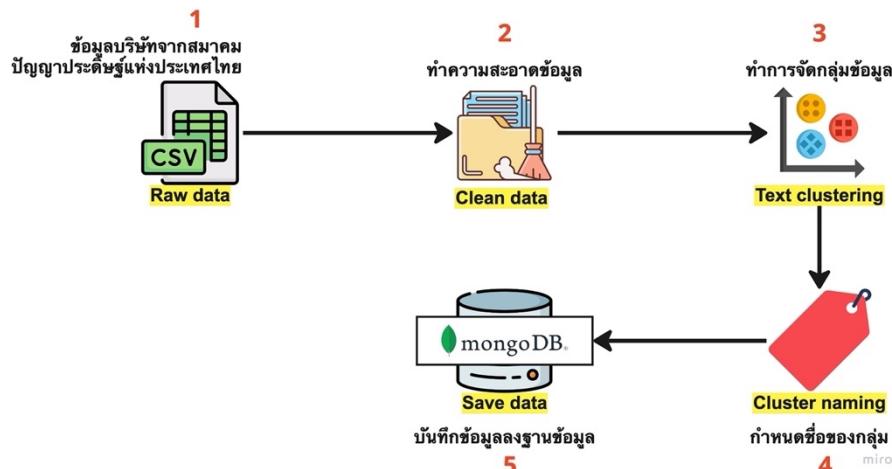
- 3.1 การเตรียมและวิเคราะห์ข้อมูล
- 3.2 การทำงานของระบบ
- 3.3 การวิเคราะห์และออกแบบระบบ
- 3.4 การออกแบบฐานข้อมูล
- 3.5 การออกแบบหน้าจอ
- 3.6 การใช้งานระบบ

#### 3.1 การเตรียมและวิเคราะห์ข้อมูล

##### ตาราง 4 การวิเคราะห์ข้อมูล

ข้อมูล	จำนวน	หน่วย
จำนวนข้อมูลสถานประกอบการทั่วหมู่ดิน	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 12 การเตรียมและวิเคราะห์ข้อมูล

จากภาพที่ 10 แสดงการเตรียมและวิเคราะห์ข้อมูลได้ดังนี้

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

### 3.1.1 การนำเข้าข้อมูลไฟล์ .csv เพื่อทำการจัดกลุ่มข้อมูล

- ดาวน์โหลดโปรเจกจาก [https://github.com/slapexs/final\\_project](https://github.com/slapexs/final_project)
- นำเข้าไฟล์ข้อมูลลงในโฟลเดอร์ data\_csv

### 3.1.2 การทำ Word segmentation

```
1 import pandas as pd
2 from pythainlp.corpus import thai_stopwords
3 from nltk.corpus import stopwords
4 from pythainlp.tokenize import word_tokenize
5 from sklearn.feature_extraction.text import TfidfVectorizer
6 import string
7 import numpy as np
```

ภาพที่ 13 การเรียกใช้ไลบรารี (Library) สำหรับคำนวณค่า TF-IDF  
จากการที่ 11 แสดงการเรียกใช้ฟังก์ชันจากไลบรารี (Library) ที่ใช้ในการคำนวณค่า TF-IDF  
ประกอบไปด้วย

1. Pandas ใช้ในการอ่านข้อมูลในไฟล์
  2. thai\_stopword เป็นรายการคำที่ไม่เลื่อนความหมายในภาษาไทย
  3. stopword เป็นรายการคำที่ไม่เลื่อนความหมายในภาษาอังกฤษ
  4. word\_tokenize เป็นฟังก์ชันที่ใช้ในการตัดคำแยกเป็นคำ ๆ จากประโยค
  5. TfidfVectorizer เป็นฟังก์ชันสำหรับคำนวณหาค่า TF-IDF จากประโยคที่ตัดคำแล้ว
  6. String เป็นคลาสของภาษา Python ที่ใช้แสดงข้อมูลตัวอักษรต่าง ๆ
  7. Numpy ใช้ในการสร้างอาเรย์สำหรับการใช้งานในการอ่านข้อมูล

ภาพที่ 14 การอ่านข้อมูลจากไฟล์และกำหนดตัวกรองการตัดคำ

จากภาพที่ 12 แสดงการอ่านข้อมูลจากไฟล์ และการกำหนดตัวกรอกในการตัดคำทั้งคำที่ไม่สื่อความหมายในภาษาไทย ภาษาอังกฤษ ตัวเลขไทย และอักขระพิเศษ พร้อมทั้งประกาศตัวแปรเพื่อเก็บค่าการจากการตัดคำ

```
1 def clean_string(detail:list) -> list:  
2     temp_clean = []  
3     for i in detail:  
4         if i not in string.punctuation and i not in string.digits and i not in spx_char and i not in th_number:  
5             temp_clean.append(i.lower())  
6     return ''.join(temp_clean)  
7
```

ภาพที่ 15 พังก์ชันสำหรับใช้ลับตัวเลข และอักขระพิเศษ

จากภาพที่ 13 แสดงการลบตัวเลข และอักษรพิเศษออก จากประยุคที่รับเข้ามาและทำการเชื่อมประยุคและคืนค่ากลับออกไป

```
● ● ●
1 def clean_stopword(token:list) -> list:
2     temp = []
3     for i in token:
4         if i not in th_stopword and i not in eng_stopword and i not in th_number:
5             temp.append(i)
6     return temp
```

ภาพที่ 16 พังก์ชันสำหรับใช้ลบคำที่ไม่สื่อความหมายและตัวเลขไทย  
จากภาพที่ 14 แสดงการลบคำที่ไม่สื่อความหมายในภาษาไทย ภาษาอังกฤษ และตัวเลขไทย  
ออกไปจากข้อมูลที่รับเข้ามาและคืนค่ากลับออกไป

```
● ● ●
1 for i in range(len(df)):
2     sample = clean_string(str(df.iloc[i]['detail']).lower())
3     text_cleaned = clean_stopword(word_tokenize(sample, None, 'newmm', False))
4     list_company_detail.append(text_cleaned)
```

ภาพที่ 17 การวนซ้ำข้อมูลเพื่อตัดคำและทำความสะอาดข้อมูล

จากภาพที่ 15 แสดงการวนซ้ำการล่งข้อมูลที่อ่านจากไฟล์เพื่อนำไปลบคำที่ไม่สื่อความหมาย  
ตัวเลข และอักษรพิเศษออกจากข้อมูลตัวแรกถึงตัวสุดท้าย และนำผลลัพธ์ไปเก็บไว้ในตัวแปร

```
● ● ●
1 def fake_tokenize(word):
2     return word
3
4 vectorizer = TfidfVectorizer(
5     analyzer='word',
6     tokenizer=fake_tokenize,
7     preprocessor=fake_tokenize,
8     token_pattern=None,
9     lowercase=True,
10    )
11 tfidf_vector = vectorizer.fit_transform(list_company_detail)
12 tfidf_array = np.array(tfidf_vector.todense())
13 df_tfidf = pd.DataFrame(tfidf_array, columns=vectorizer.get_feature_names_out())
14 df_tfidf = df_tfidf.drop(df_tfidf.columns[[k for k in range(-15, 0, 1)]], axis = 1)
15 print(df_tfidf)
```

ภาพที่ 18 การเทrnและทำการทดสอบโมเดลการคำนวณค่า TF-IDF

จากภาพที่ 16 แสดงการสร้างเวกเตอร์ของการคำนวณค่า TF-IDF การเทrnข้อมูล และการทดสอบการประมาณผลจากโมเดลที่เทrnประกอบไปด้วย

1. vectorizer เป็นการ tren โมเดลสำหรับการคำนวณค่า TF-IDF
2. tfidf\_vector เป็นการทดสอบและสร้างเวกเตอร์ของคำในแต่ละประโยค
3. tfidf\_array เป็นการนำเวกเตอร์มาเปลี่ยนให้อยู่ในรูปแบบของอาเรย์
4. df\_tfidf เป็นการนำข้อมูลในอาเรย์มาสร้างเป็นตารางข้อมูล

```

python tfidf.py
[1643 rows x 8526 columns]

```

ภาพที่ 19 ผลลัพธ์การคำนวณค่า TF-IDF

จากภาพที่ 17 แสดงผลลัพธ์น้ำหนักแต่ละคำจากที่ผ่านการคำนวณค่า TF-IDF และแสดงออกมาเป็นตารางข้อมูลเรียงลำดับข้อมูลตั้งแต่ประโยคแรกถึงสุดท้าย

### 3.1.3 การจัดกลุ่มข้อมูล (Clustering)



```

1 from sklearn.cluster import KMeans
2 from sklearn.decomposition import PCA
3 from sklearn.preprocessing import StandardScaler

```

ภาพที่ 20 การเรียกใช้ไลบรารีสำหรับการจัดกลุ่มข้อมูลด้วยเคมีน์ (K-Means) จากภาพที่ 18 แสดงการนำเข้าไลบรารีจาก Scikitlearn เพื่อทำการจัดกลุ่มข้อมูลโดยมีรายละเอียดดังนี้

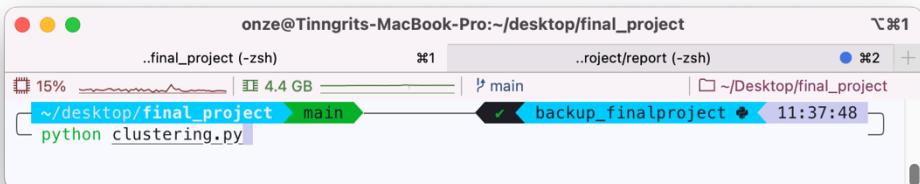
1. KMeans เป็นการเรียกใช้ลักษณะที่มีสำหรับการจัดกลุ่ม
2. PCA (Principle Components Analysis) ซึ่งเป็นวิธีการลด dimension ของ Feature ลงช่วยลดทอนความซ้ำซ้อนของข้อมูลทำให้เรา train model ได้ง่ายขึ้น
3. StandardScaler เป็นตัวแปลงค่าตัวเลขให้อยู่ในปริมาณที่ใกล้เคียงกัน

```

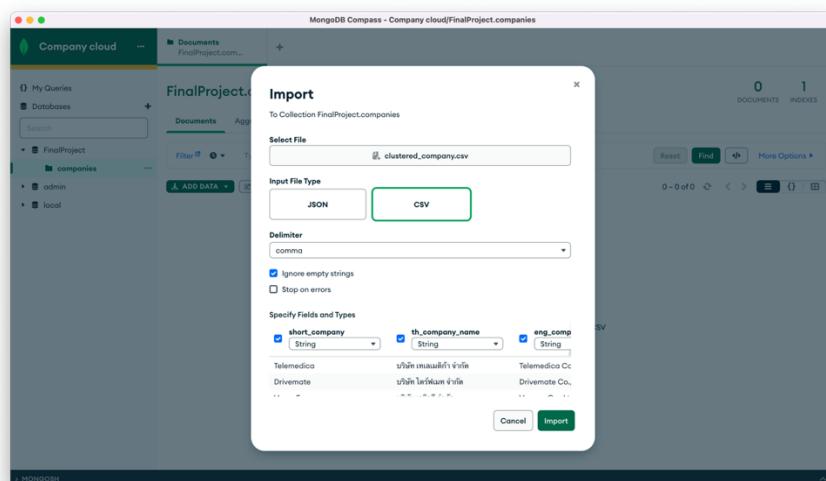
1 k = 7
2 kmeans = KMeans(n_clusters=k, random_state=1)
3 # Fit model
4 kmeans.fit(df_tfidf[['x_value', 'y_value']])
5 clusters = kmeans.labels_

```

- ภาพที่ 21 แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล
- เมื่อกำหนดค่าการจัดกลุ่มเรียบร้อยเรียกใช้ไฟล์ clustering.py ใน Terminal เพื่อทำการจัดกลุ่มและบันทึกผลลัพธ์

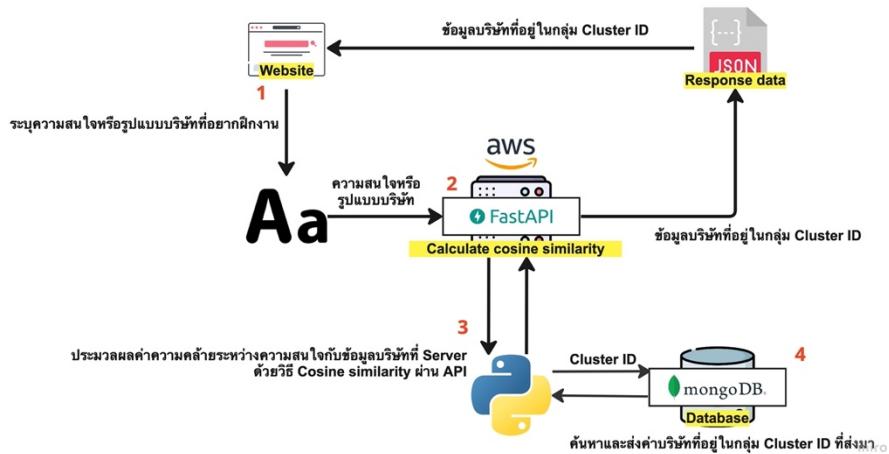


- ภาพที่ 22 การเรียกใช้งานไฟล์ clustering.py เพื่อจัดกลุ่มข้อมูลและบันทึกผลลัพธ์
- ### 3.1.3 การนำเข้าข้อมูล Clustering เข้าสู่ฐานข้อมูล
- นำเข้าข้อมูลลงฐานข้อมูล MongoDB โดยโปรแกรม MongoDB Compass
  - เชื่อมต่อ MongoDB Compass กับ Mongodb Atlas
  - นำเข้าข้อมูลด้วยไฟล์ .CSV ที่เป็นผลลัพธ์จากการจัดกลุ่มข้อมูล



ภาพที่ 23 การนำเข้าข้อมูลลงสู่ฐานข้อมูล MongoDB

### 3.2 การทำงานของระบบ



ภาพที่ 24 การทำงานของระบบ

จากภาพที่ 11 แสดงการทำงานของระบบได้ดังนี้

- Website ใช้ระบุความสนใจเพื่อส่งค่าไปประมวลผลความคล้ายคลึงกับข้อมูลบริษัท
- Server ใช้ประมวลผลความคล้ายคลึงกันระหว่างความสนใจที่ได้รับมาและข้อมูลบริษัทที่อยู่ในฐานข้อมูลด้วยเทคนิค Cosine similarity โดยภาษา Python และส่งค่ากลับไปเป็น Cluster ID
- เมื่อได้ Cluster ID แล้วนำไปค้นหาบริษัทที่ Cluster ID ตรงกันในฐานข้อมูลและคืนค่า Response API เป็นข้อมูลในรูปแบบ JSON ที่มีข้อมูลบริษัทที่อยู่ใน Cluster ID นั้น
- mongoDB เป็นฐานข้อมูลที่เก็บข้อมูลบริษัทไว้ และรอให้เซิร์ฟเวอร์เรียกใช้ข้อมูลเพื่อนำไปแสดงผล

### 3.3 การวิเคราะห์และออกแบบระบบ

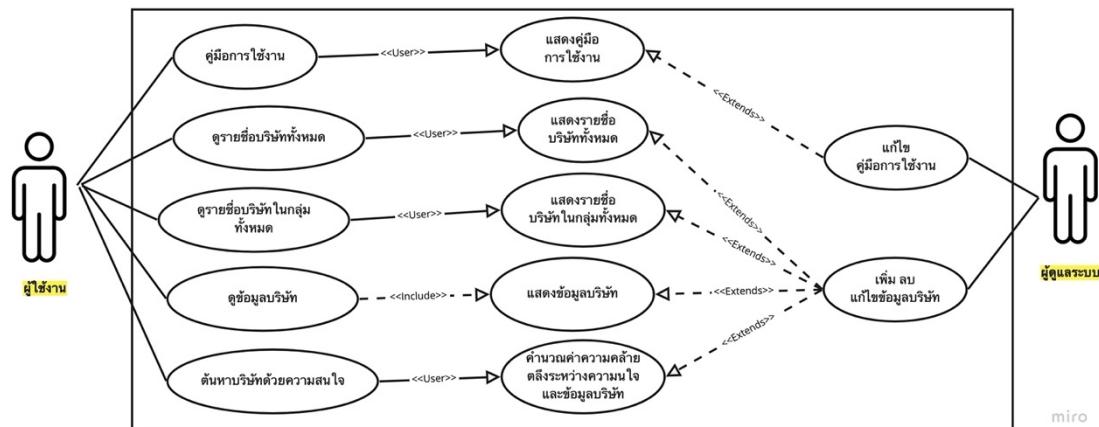
#### 3.3.1 การวิเคราะห์ระบบ

การวิเคราะห์ระบบและการออกแบบ (System Analysis and Design) คือ วิธีการที่ใช้ในการสร้างระบบสารสนเทศขึ้นมาใหม่ในธุรกิจ ได้ธุรกิจหนึ่งหรือระบบบ่อยของธุรกิจ นอกจากการสร้างระบบสารสนเทศใหม่แล้ว การวิเคราะห์ระบบ ช่วยในการแก้ไขระบบสารสนเทศเดิมที่มีอยู่แล้วให้ดีขึ้น การวิเคราะห์ระบบ คือ การหาความต้องการ (Requirements) ของระบบสารสนเทศ คือ อะไร หรือต้องการเพิ่มเติมอะไรเข้ามาในระบบ การออกแบบ คือ การนำความต้องการของระบบมาเป็นแบบแผน หรือเรียกว่าพิมพ์เขียวในการสร้างระบบสารสนเทศนั้นให้ใช้งานได้จริง

#### 3.3.2 ผู้ใช้และแบบแกรม (Use Case Diagram)

ผู้ใช้และแบบแกรม (Use Case Diagram) คือ แผนภาพที่ใช้แสดงปฏิสัมพันธ์ระหว่างระบบงานและลิงก์ที่อยู่ในระบบงาน และแสดงให้เห็นถึงส่วนประกอบทั้งหมด หรือ ภาพรวมของระบบ เป็นรากฐานในการเริ่มต้นการวิเคราะห์ระบบ โดยคนหาคำว่าระบบทำอะไร โดยไม่สนใจกับการทำงานหรือเทคนิคการทำางานเปรียบเสมือน "กล่องดำ" โดย Use Case Diagram จะช่วยให้

ผู้พัฒนาระบบสามารถแยกแยะกิจกรรมที่อาจจะเกิดขึ้นในระบบ เป็น Diagram พื้นฐาน ที่สามารถอธิบายสิ่งต่าง ๆ ได้โดยใช้รูปภาพที่ไม่ซับซ้อน



ภาพที่ 25 Use Case Diagram ระบบแนะนำสถานที่ทำงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเค้มีนท์ (K-Means)

#### 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:	กรณีที่แสดงผลจากการคนหาด้วยความสนใจ 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"> <li>ผู้ใช้งานพิมพ์ความสนใจหรือรูปแบบธุรกิจที่ช่องคนหาและ</li> <li>แสดงรายชื่อบริษัทที่อยู่ในกลุ่มที่ระบบแนะนำ</li> <li>ผู้ใช้คลิกเลือกบริษัทเพื่อดูข้อมูลบริษัท</li> </ol>
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"> <li>ผู้ดูแลระบบแก้ไขข้อมูลคู่มือการใช้งาน</li> <li>Deploy เพื่ออัปเดตระบบ</li> </ol>
Alternate/Exceptional Flow:	None

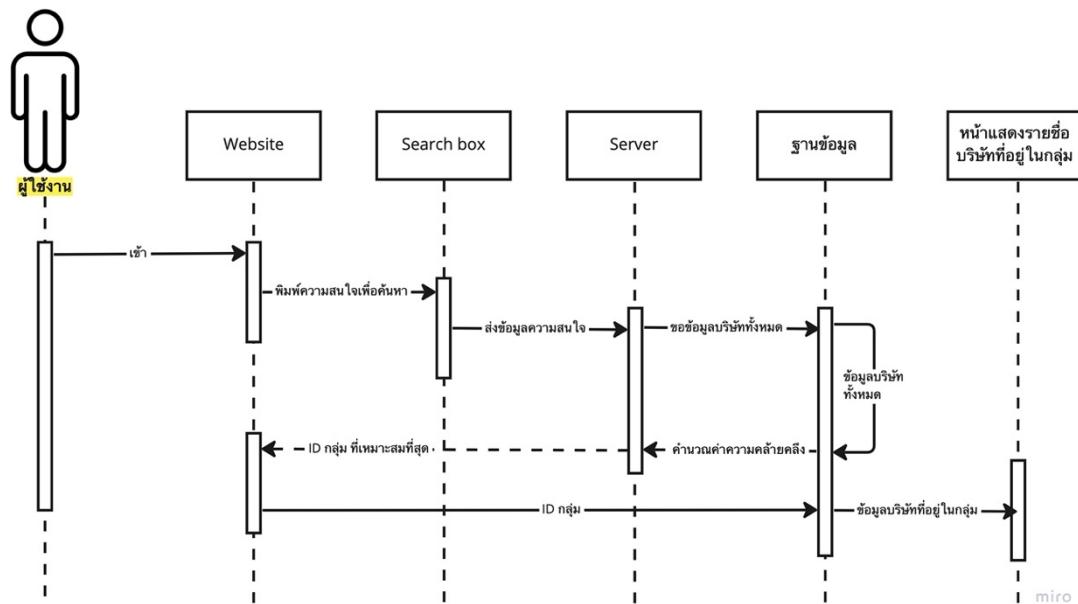
ตาราง 11 คำอธิบาย Use case เพิ่ม ลบ แก้ไขข้อมูลบริษัท

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

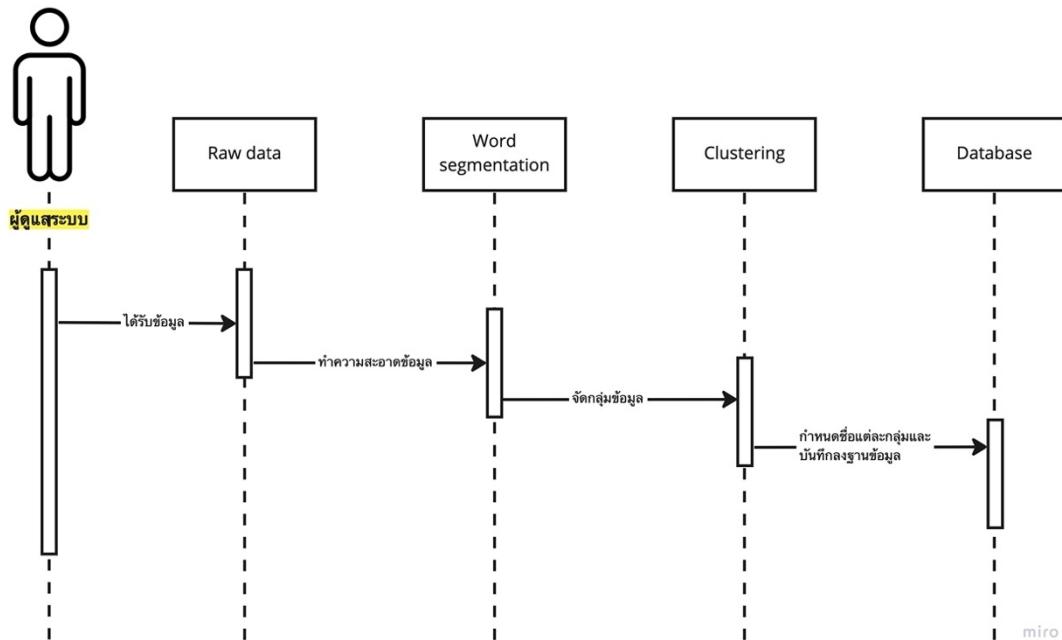
แสดงลำดับเวลา และเส้นที่ใช้เพื่อแสดงกิจกรรมที่เกิดขึ้นจากคลาสรหีอวัตถุในแผนผังการทำงานภายใน Sequence Diagram จะใช้สีเหลี่ยมแทนสเมือนคลาสและวัตถุโดยภายในจะมีชื่อของคลาสหรือวัตถุประกอบอยู่ในรูปแบบ [Object]: Class



ภาพที่ 26 Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้

ตาราง 12 อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้

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



ภาพที่ 27 Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบัญชีใหม่

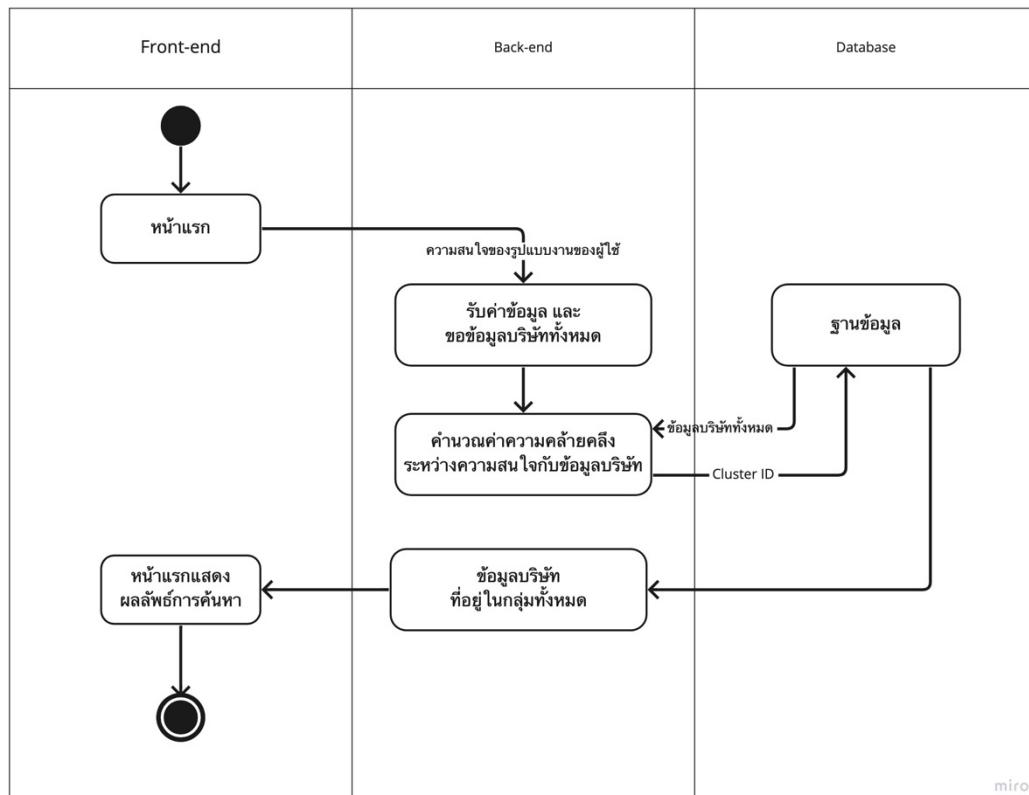
ตาราง 13 อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบัญชีใหม่

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

### 3.3.4 แอคทิวิตี้ไดอะแกรม (Activity Diagram)

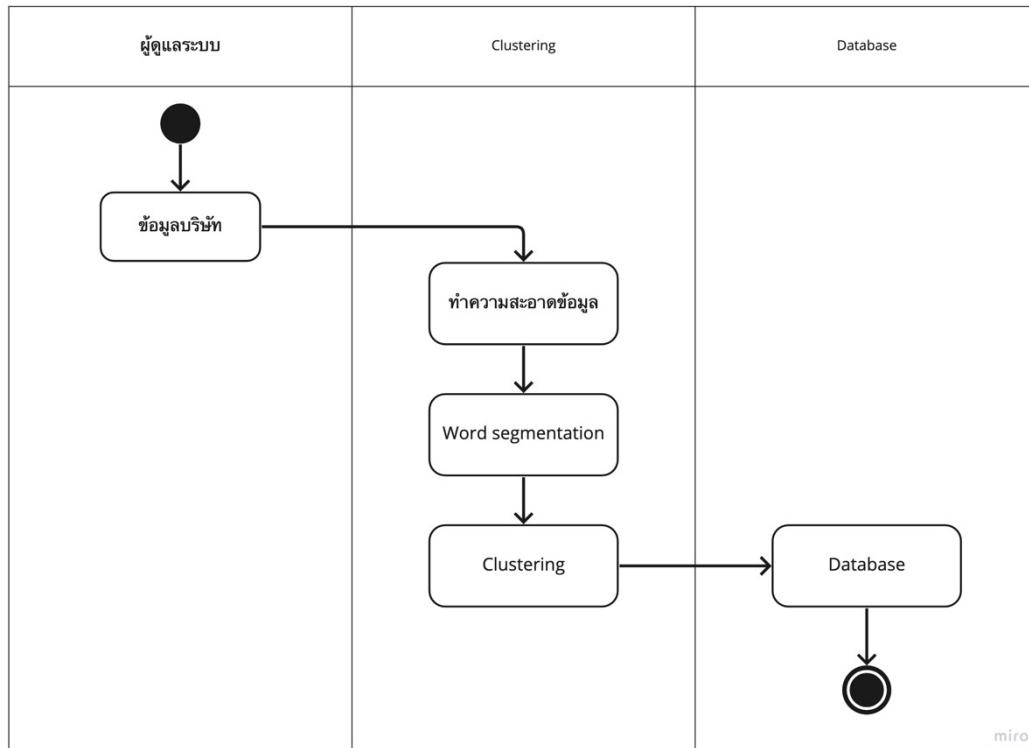
Activity Diagram หรือแผนภาพกิจกรรม ใช้อธิบายกิจกรรมที่เกิดขึ้นในลักษณะกราฟแสดงการให้ผลของการทำงาน (Workflow) จะมีลักษณะเดียวกับ Flowchart โดยขั้นตอนในการทำงานแต่ละขั้นจะเรียกว่า Activity ดังภาพที่ 16 และภาพที่ 17 มีรายละเอียดดังนี้ การใช้งาน Activity Diagram

1. อธิบายกราฟแสดงการให้ผลของการทำงาน (Workflow)
2. แสดงขั้นตอนการทำงานของระบบ



ภาพที่ 28 Activity Diagram ของผู้ใช้งาน

จากราฟที่ 26 แสดงขั้นตอนการทำงานของระบบเมื่อผู้ใช้คนหากข้อมูลด้วยความสนใจของผู้ใช้งานจะเริ่มต้นจากการที่ผู้ใช้ระบุความสนใจ จากนั้นทำการส่งข้อมูลไปคำนวณค่าความคล้ายคลึงผ่าน API และเรียกข้อมูลบริษัทที่มีความคล้ายมากที่สุดมาแสดงผลหน้าเว็บไซต์



ภาพที่ 29 Activity Diagram ของผู้ดูแลระบบ

จากภาพที่ 27 แสดงการทำงานของการจัดกลุ่มข้อมูลเพื่อนำไปใช้ในระบบโดยการทำงานเริ่มต้นที่นำข้อมูลที่ได้เก็บรวบรวมลงในไฟล์จากนั้นทำการทำ Word segmentation และทำการจัดกลุ่มข้อมูล สุดท้ายบันทึกข้อมูลเพื่อนำไปนำเข้าลงฐานข้อมูล MongoDB

### 3.4 การออกแบบฐานข้อมูล

#### 3.4.1 ER Diagram

ER Diagram คือ แบบจำลองที่ใช้อธิบายโครงสร้างของฐานข้อมูลซึ่งเขียนออกมาในลักษณะของรูปภาพ การอธิบายโครงสร้างและความสัมพันธ์ของข้อมูล (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 เพื่อให้ลื่อสารอย่างตรงกัน

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 30 ER Diagram ระบบແນ່ນຳບັນທຶກສໍາຮັບຜິດການຕາມຄວາມສນິໃຈ  
ດ້ວຍເຕັກໂນໂລຢີປັນຍາປະຕິເມືອງ

#### 3.4.2 ພຈນານຸກຮມຂອ່ມູນ (Data Dictionary)

ໜັງຈາກທີ່ວິເຄາະທີ່ຮະບບແລ້ວ ຜູ້ສຶກສາໄດ້ອອກແບບຈຸານຂອ່ມູນ ໂດຍອອກແບບ  
ໂຄຮງສ້າງຂອງຮະບບ ຊື່ປະກອບໄປດ້ວຍຕາຮາງ ຈຳນວນ 1 ຕາຮາງ ແລະໄດ້ອີ່ນບາຍ ຊື່ອ ຕາຮາງ(File  
Name), ຄໍາອີ່ນບາຍ(Description), ຊື່ຂໍ້ອ່ມູນ(Field Name), ຜົນດີຂອງຂໍ້ອ່ມູນ(Type), ຂາດທີ່ເກັບ(Length),  
ລັກຂະນະທີ່ເກັບຄາ(Format), ຜົນດີຂອງຄີ່ຍ(Key) ດັ່ງຕົວໄປນີ້

ຕາຮາງ 14 ພຈນານຸກຮມຂອ່ມູນບັນທຶກ

File name: companies					
Description: ຕາຮາງເກັບຂໍ້ອ່ມູນບັນທຶກທີ່ໜ້າ					
Field name	Type	Length	Format	Description	Key
_id	String	50	ຕົວອັກຂຽນ	ຮັດສບບັນທຶກ	Primary key
short_company	String	50	ຕົວອັກຂຽນ	ຊື່ອົບບັນທຶກ	Null
th_company_name	String	50	ຕົວອັກຂຽນ	ຊື່ບັນທຶກ ພາສາໄທ	Null
eng_company_name	String	50	ຕົວອັກຂຽນ	ຊື່ບັນທຶກ ພາສາອັກຄູນ	Null
type_business	String	50	ຕົວອັກຂຽນ	ປະເກດ ຫຼຸງກິຈ	Null
product	String	255	ຕົວອັກຂຽນ	ປະເກດ ສິນຄາ	Null
type_innovation	String	50	ຕົວອັກຂຽນ	ປະເກດ ເຕັກໂນໂລຢີ	Null
detail	String	255	ຕົວອັກຂຽນ	ຮາຍລະເຂົ້າດ ຫຼຸງກິຈ	Null
owner	String	255	ຕົວອັກຂຽນ	ເຈົາຂອງ	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 ซึ่งจะมีการออกแบบในทางด้านของหน้าตา ปุ่ม ช่องที่ใช้พิมพ์สำหรับคนหาข้อมูล ตัวอักษร สี และรูปภาพเป็นต้น

การออกแบบหน้าจอสำหรับการค้นหาสถานประกอบการสำหรับฝีมือด้วยเทคโนโลยีปัญญาประดิษฐ์นั้น จะเน้นในแพลตฟอร์มที่เป็นหน้าจอคอมพิวเตอร์หรือ Desktop เป็นหลัก โดยมีรายละเอียดดังนี้

#### 3.5.1 หน้าแรก



1 Intern Assistant

รายการสถานประกอบการทั้งหมด หน่วยงานสถานประกอบการ ▾

2

ค้นหาสถานประกอบการ  
ระบุรายละเอียดความสนใจที่คุณต้องการ

3

สถานประกอบการ บ้าสันดิ

<b>Ernsen - Schmeler</b> 192 Screeching Street, West Malling, BD72 6RG Joe@toppid.org 07183 776-242 Online marketing	<b>Luettgen - Hintz</b> 400 Kindly Road, Llandover, W27 1TF Diarumud@fuel.org 07223 538-213 Online marketing	<b>Stehr, Doyle and Schultz</b> 430 Payment Avenue, Surrey, TS77 7HA Harley@wealthy.co 07154 387-953 Network
<b>Lemke and Sons</b> 467 Chilly Road, Winsford, TA88 2SR	<b>Champlin - Kessler</b> 229 Pencil Close, Countess Wear, BN78 BRH	<b>Christiansen Group</b> 460 Annoying Crescent, Newton Stewart, PH6 6S

ภาพที่ 31 หน้าแรก

จากการที่ 29 แสดงการออกแบบหน้าแรกของเว็บไซต์ซึ่งประกอบไปด้วยส่วนประกอบดังนี้ หมายเลขอ 1 เมนูหลักสามารถเชื่อมโยงไปยังหน้าแสดงรายละเอียดของเว็บไซต์ หน้าสถานประกอบการทั้งหมด และหน้าสถานประกอบการแต่ละประเภท

หมายเลขอ 2 เป็นช่องสำหรับระบุความสนใจของผู้ใช้เพื่อนำไปค้นหาผลลัพธ์ที่มีความคล้ายหมายเลขอ 3 เป็นการแนะนำบริษัทที่น่าสนใจจากที่อยู่ในฐานข้อมูล

### 3.5.2 หน้าเกี่ยวกับ

The screenshot shows the 'About' page of a web application. At the top, there is a navigation bar with icons for Intern Assistant, Home, About the system, and Contact. Below the navigation bar, there is a section titled 'เกี่ยวกับ' (About) with a red circle containing the number 1. This section contains a brief text about the system's purpose and a flow diagram. The flow diagram shows a CSV file being processed by a central node (a grid with colored circles) to produce a database output. Below the diagram, there is a list of service links: API, Github, Document, and Article.

### ภาพที่ 32 หน้าเกี่ยวกับ

หมายเลขอ 1 คือการอธิบายรายละเอียดเกี่ยวกับระบบและรูปแบบการทำงานในส่วนต่าง ๆ

### 3.5.3 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

The screenshot shows the 'List of Companies' page of a web application. At the top, there is a navigation bar with icons for Intern Assistant, Home, About the system, and Contact. Below the navigation bar, there is a section titled 'หน้าเดียว: Online marketing' with a red circle containing the number 1. This section contains a table listing various companies. The table has two columns: 'Name' and 'Description'. The companies listed include: Von, Kunde And Stracke, Bode And Sons, Grady, Huels And Runte, Balistreri Inc, Cartwright, Goyette And Watsica, Schulist LLC, Olson LLC, Davis, Dibert And Schuster, Ferry - Howe, Konopelski Group, Rutherford - Denesik, Grant And Sons, Wilkinson - Hegmann, Conn - Bergstrom, Ernsler - Schmeler, Tromp LLC, Gusikowski - Considine, Dicki, Welch And Rippin, Boyer - Daugherty, Gutkowski - Kautzer, Ledner - Pfeffer, Mohr, Herzog And Terry, Harris, Veum And Kertzmann, Rolfson - Wiza, Baumbach - Raynor, Dach, Cummings And Lindgren, Halvorson - Nikolaus, Mueller - Hodkiewicz, and Conn - Bergstrom.

### ภาพที่ 33 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หมายเลขอ 1 แสดงรายชื่อบริษัทที่ที่อยู่ในกลุ่มที่เลือกดูทั้งหมดและสามารถคลิกเพื่ออ่านรายละเอียด

### 3.5.4 หน้าแสดงรายชื่อบริษัททั้งหมด

หน้าแสดงรายชื่อบริษัททั้งหมด

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

ภาพที่ 34 หน้าแสดงรายชื่อบริษัททั้งหมด  
หมายเลข 1 รายชื่อบริษัททั้งหมดที่อยู่ในฐานข้อมูลและสามารถคลิกดูข้อมูลบริษัทด้วย

### 3.5.5 หน้าแสดงผลลัพธ์รายชื่อบริษัท

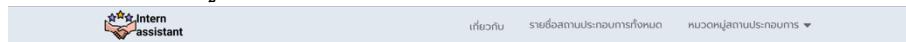
หน้าแสดงผลลัพธ์รายชื่อบริษัท

Ernsler - Schmeler  
Tromp LLC  
Gusikowski - Considine  
Dicki, Welch And Rippin  
Boyer - Daugherty  
Von, Kunde And Stracke  
Bode And Sons  
Grady, Huels And Runte  
Balistreri Inc  
Cartwright, Goyette And Watsica

ภาพที่ 35 หน้าแสดงผลลัพธ์รายชื่อบริษัท

หมายเลข 1 ซองสำหรับระบุความสนใจของผู้ใช้เพื่อค้นหาบริษัทที่คล้ายคลึง  
หมายเลข 2 ผลลัพธ์รายชื่อบริษัทที่มีความคล้ายคลึงกับความสนใจที่ผู้ใช้ระบุ

### 3.5.6 หน้าแสดงข้อมูลบริษัท



- 1 **Ernsler - Schmeler**  
Online marketing
- 2 467 Chilly Road, Winsford, TA88 2SR  
Joe@torpid.org  
07183 776 242
- 3 Repellendus laudantium dignissimos deleniti. Officiis et maiores quod veritatis dignissimos voluptatem possimus. Magni tempore sed. Animis eum voluptas dolorum esse amet quisquam. Tempore suscipit animi harum voluptatem impedit. Temporibus sed aperiam impedit cum modi. Autem architecto est eveniet Et dolores assumenda numquam qui qui. Debitis et perspicatis ad iste. Minima repudiandae dolor rerum et aut. Atque magni ullam assumenda a consecutetur molestias tenetur dolorem eveniet. Et qui ea quam ea quia. Expedita libero enim ut. Vitae iusto sed molestiae ut optio dolor perferendis perferendis. Praesentium ipsum provident qui ut error beatae quibusdam. Et qui animi qui in voluptas sed.



หมายเหตุ 1 แสดงชื่อของบริษัท

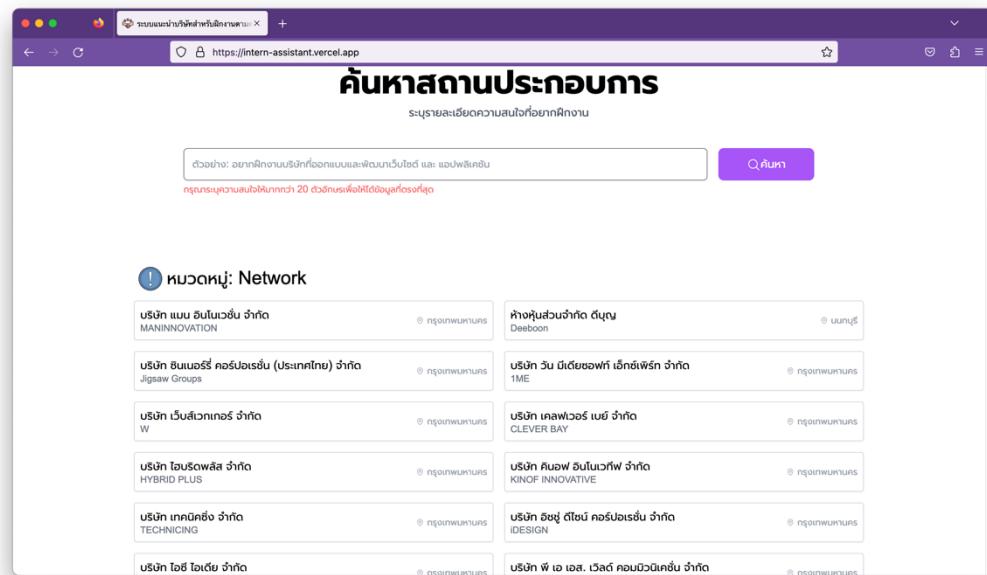
หมายเหตุ 2 แสดงรายละเอียดข้อมูลที่อยู่ ข้อมูลการติดต่อของบริษัท

หมายเหตุ 3 แสดงข้อมูลรายละเอียดรูปแบบธุรกิจที่บริษัทดำเนินกิจการอยู่

## 3.6 การใช้งานระบบ

1. ระบุความสนใจในรูปแบบธุรกิจ หรือสิ่งที่อยากรับในการฝึกงานลงในช่องค้นหา
2. เมื่อได้ผลลัพธ์บริษัทที่มีความคล้ายคลึงกับสิ่งที่ค้นหาแล้วผู้ใช้สามารถทำการพิจารณา

บริษัทเพื่อเลือกตัดสินใจในการฝึกงานได้



ภาพที่ 37 ตัวอย่างผลลัพธ์จากการค้นหาด้วยความสนใจของผู้ใช้

## บทที่ 4

### ผลการดำเนินงาน

การทำงานของระบบแนะนำสำหรับผู้ใช้งานตามความสนใจด้วยเทคโนโลยีปัญญาประดิษฐ์ 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 การกำหนดคำที่ไม่สืบความหมายหรือ Stop word เช่น ที่ นี่ อีน เป็นต้น ออกไปจากประโยคเพื่อให้ได้ประโยคที่มีเนื้อหาใจความดีที่สุด และอีกขั้นตอนสำคัญคือการทำคำอ่านภาษาไทย ที่จะได้ทราบว่าประโยคนั้น ๆ กำลังสื่อถึงเรื่องไหนเป็นสำคัญด้วยเทคนิค TF-IDF ที่เป็นการทำหน้าหนักของคำนั้น ๆ ในประโยค

ในการทำงานประมวลผลเกี่ยวกับการประมวลผลภาษาธรรมชาติจำเป็นต้องมีการทำตัดคำออกเป็นคำ ๆ เพื่อจะได้่ายและนำไปเข้าสู่กระบวนการต่าง ๆ ได้อย่างง่ายโดยปกติแล้วการทำคำในภาษาอังกฤษนั้นสามารถตัดได้โดยใช้การเว้นวรรคเป็นเงื่อนไขในการตัด แต่ในภาษาไทยนั้นการทำเช่นนั้นไม่ได้มีการเว้นวรรคคำเหมือนภาษาอังกฤษทำให้การทำตัดคำนั้นจะใช้เว้นวรรคมาตัดคำตลอดไม่ได้ จำเป็นต้องใช้อัลกอริทึมอื่น ๆ เช่นมาช่วย เช่นการใช้ Dictionary-based, Maximum Matching เป็นต้น

และในภาษา Python ก็มีเครื่องมือที่สามารถช่วยอำนวยความสะดวกในการตัดคำภาษาไทยอย่าง Pythainlp ที่ผู้วิจัยได้เลือกใช้ในโครงการนี้ในไลบรารีนั้นสามารถตัดคำได้หลาย Engine ด้วยกันและในแต่ละตัวเลือกก็ใช้อัลกอริทึมต่างกันยกตัวอย่างเช่น 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

### ภาพที่ 38 ตัวอย่างข้อมูลต้นฉบับ

ภาพที่ 39 ตัวอย่างการตัดคำโดยใช้ Engine newmm

ภาพที่ 40 ตัวอย่างการตัดคำโดยใช้ Engine longest

ภาพที่ 41 ตัวอย่างการตัดคำโดยใช้ Engine deepcut

### ผลการทดสอบความแม่นยำการตัดคำ

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

จากการทดสอบพบว่า Engine newmm ในไลบรารี Pythainlp นั้นมีความแม่นยำกับที่ผู้วิจัยตัดมากที่สุดที่ 90.99% รองลงมาที่ longest คาดความแม่นยำอยู่ที่ 83.04% และความแม่นยำน้อยที่สุดคือ deepcut ที่ความแม่นยำ 76.65% ตั้งภาพที่ 29 ดังนั้นโครงการนี้จึงใช้ Engine newmm ในการตัดคำเพื่อนำไปประมวลผลต่อในขั้นตอนอื่น ๆ

```
onze@Tinngrits-MacBook-Pro:~/Documents/final_project
..final_project (-zsh) ❶ ..final_project (-zsh) ❷ ~D/final_project
└─ python calculate_accuracy.py
newmm: 90.99%
deepcut: 76.65%
longest: 83.04%
~/Documents/final_project report +1 !2 ?2 ❸ 5s 14:41:13
```

ภาพที่ 42 ผลการวัดคาดความแม่นยำในการตัดคำของ Engine ในไลบรารี Pythainlp

#### 4.1.2 การคำนวณค่า TF-IDF

เมื่อได้เครื่องมือที่จะช่วยตัดคำแล้วขั้นตอนต่อไปคือการนำมาทำการจัดกลุ่มข้อมูลโดยเทคนิคที่เลือกใช้คือการค่า TF-IDF (Term Frequency–Inverse Document Frequency) เพื่อหาว่าคำไหนในประโยคนั้นเป็นคำสำคัญของประโยคนั้น ๆ โดยการวัดจากน้ำหนักของคำด้วยวิธีดังกล่าว ตั้งภาพที่ 30 ที่แสดงการตัดคำและเรียงจากน้ำหนักของคำ 5 อันดับแรก

```
onze@Tinngrits-MacBook-Pro:~/documents/final_project
~/documents/final_project tfidf
python main.py
[('based', 'datacenter', 'solutions', 'map', 'embedded')]
list(['printed', 'พิมพ์', 'พิมพ์', 'แผนภูมิ', 'แผนภูมิ'])
list(['อินดี้', 'สอน', 'วิเคราะห์', 'สอน', 'ตั้งเป้าหมาย'])
list(['as', 'a', 'service', 'open', 'api'])
list(['ไฟล์', 'สอนໄน์', 'ร่าง', 'สอนเบนช์', 'ต้น'])
list(['เทคโนโลยีสารสนเทศ', 'สร้าง', 'គរបាល', 'สอนแบบ', 'ระบบ'])
list(['เอกสาร', 'n', 'อินเด็กซ์', 'docflow', 'icr'])
list(['blockchain', 'technology', 'เทคโนโลยีสารสนเทศ', 'สร้าง', 'ให้คำปรึกษา'])
list(['app', 'dolphin', 'finpoint', 'know', 'ตัว'])
list(['renthub', 'เอกสาร', 'สอน', 'สอน', 'th'])
```

ภาพที่ 43 ตัวอย่างการตัดคำและเรียงคำที่มีน้ำหนักมากที่สุด 5 อันดับ

	aaa	ab	abap	abeam	ablerex	abroad	academic	acceptance	access	accessories	...	ໄອທີ	ໄອສ ຕຽນ	ໄອ ເຕືອ	ໄອ ແພດ	ໄອ ໂນບາຍ	ໄອ ໂກ	ໄອ ໂກໂຄຣເຈນ
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 44 ตัวอย่างตาราง TF-IDF แสดงน้ำหนักของคำ

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

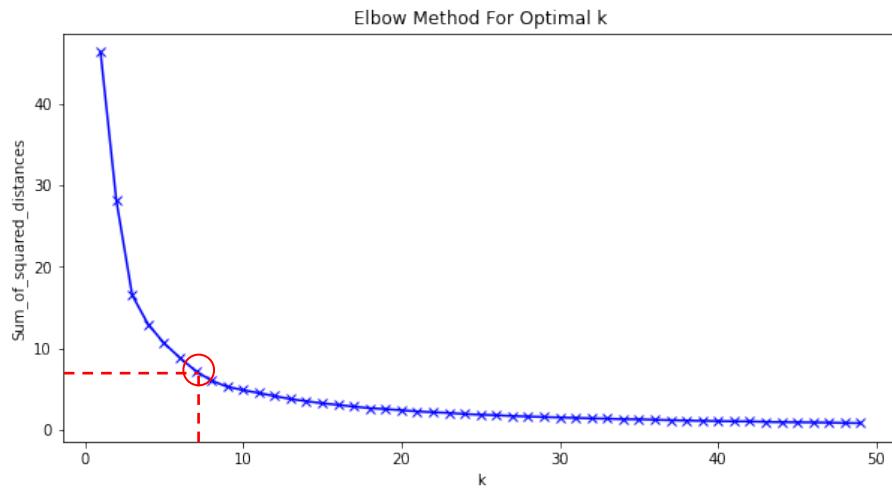
ประโยชน์เริ่มต้น: บริการ Platform รวมสถานที่จัดประชุมและผู้ให้บริการสำหรับงานอีเว้นท์ที่ดีที่สุด คำที่ไม่สื่อความหมาย (Stop word): รวมจัดและที่ที่สุด

ประโยชน์ใหม่: บริการPlatform/สถานที่ประชุมผู้ให้บริการสำหรับงานอีเว้นท์ดี

```
onze@Tinngrits-MacBook-Pro:~/Documents/final_project
..final_project (~zsh)
..final_project (~zsh)
python report/demo/main.py
Default: บริการ Platform รวมสถานที่จัดประชุมและผู้ให้บริการสำหรับงานอีเว้นท์ที่ดีที่สุด
Keep stop word: ['บิ๊กการ', 'Platform', 'รวม', 'สถานที่', 'จัด', 'ประชุม', 'และ', 'ผู้ให้บริการ', 'สำหรับ', 'งาน', 'อีเว้นท์', 'ที่', 'ดี', 'ที่สุด']
Removed stop word: ['บิ๊กการ', 'platform', 'สถานที่', 'ประชุม', 'ผู้ให้บริการ', 'สำหรับ', 'งาน', 'อีเว้นท์', 'ที่']
```

ภาพที่ 45 ตัวอย่างการตัดคำและลบ Stop word

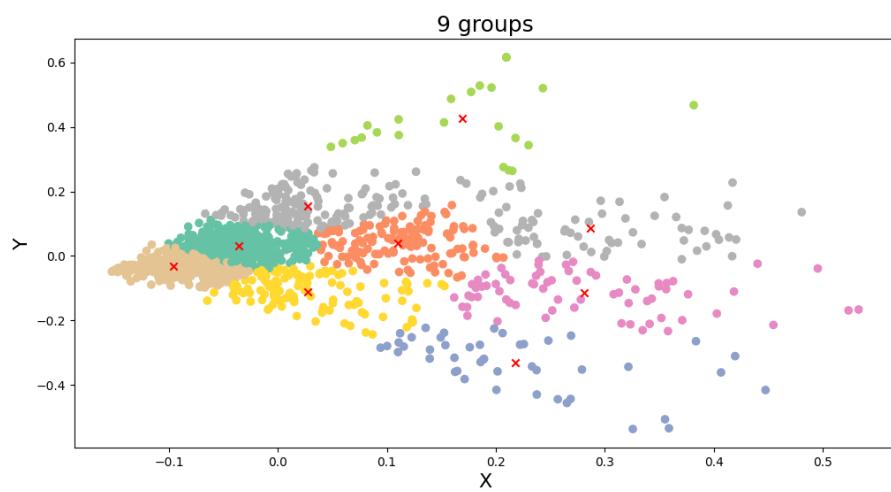
การจัดกลุ่มของข้อมูลหรือการทำ Text clustering นั้นโครงงานนี้จะใช้เทคนิค K-Means มาใช้ในการจัดกลุ่มโดยเลือกจำนวนกลุ่มจากการทำ Elbow method เพื่อหาจำนวนกลุ่มที่ดีที่สุดได้ดังภาพที่ 32 โดยจำนวนกลุ่มที่เหมาะสมจะอยู่ที่บริเวณส่วนโคงคล้ายของศอกในที่นี่จะประมาณกลุ่มได้ 6-9 กลุ่ม



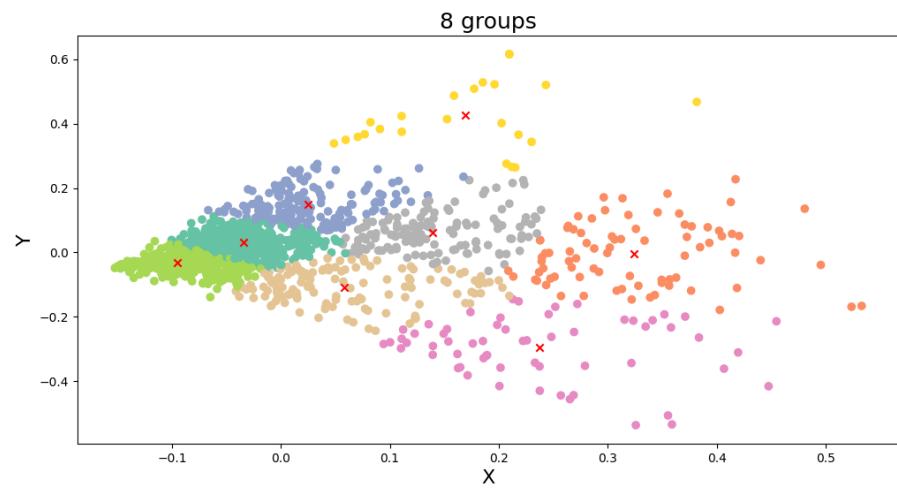
ภาพที่ 46 การทำ Elbow method

จากนั้นทำการทดลองจัดกลุ่มข้อมูลด้วยจำนวนกลุ่มที่แตกต่างกันแต่ข้อมูลเดียวกัน ทำการทดลองจัดกลุ่มจำนวน 4 กลุ่มได้ดังนี้

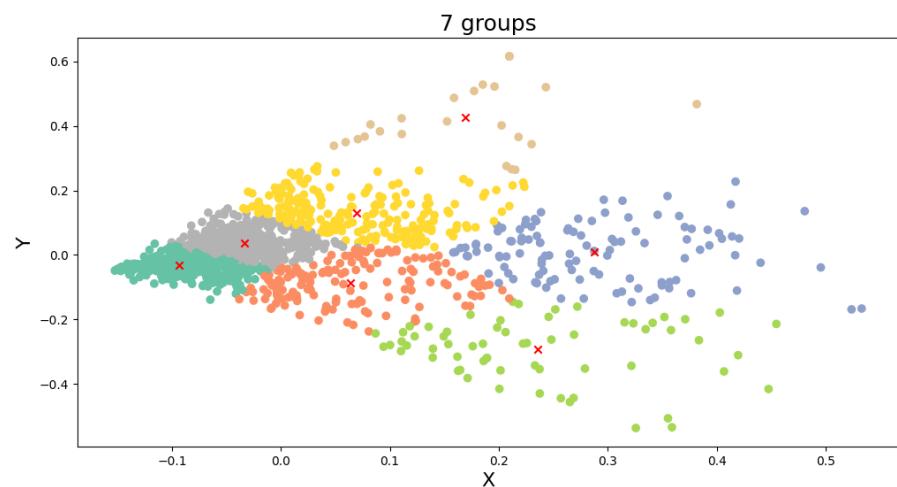
1. จำนวน 9 กลุ่ม พบร่วมของเขตของข้อมูลค่อนข้างแคบมากและมีกลุ่มที่มีเนื้อหาซ้ำกันมากกว่า 1 กลุ่ม
2. จำนวน 8 กลุ่ม พบร่วมของเขตของข้อมูลค่อนข้างแคบและเนื้อหาบางส่วนจะปนอยู่ในกลุ่มอื่นๆที่มีใกล้กัน
3. จำนวน 7 กลุ่ม พบร่วมของเขตของข้อมูลค่อนข้างตีและเนื้อหาในกลุ่มนั้นมีประปนกันอย่างมากและไม่มีกลุ่มที่มีเนื้อหาซ้ำกัน
4. จำนวน 6 กลุ่ม พบร่วมของเขตของข้อมูลกว้างมากและทำให้ใน 1 กลุ่มนั้นมีเนื้อหาที่มากกว่า 1 อย่างทำให้ไม่สามารถระบุแหล่งได้ว่าเกี่ยวกับเรื่องใดเป็นหลัก



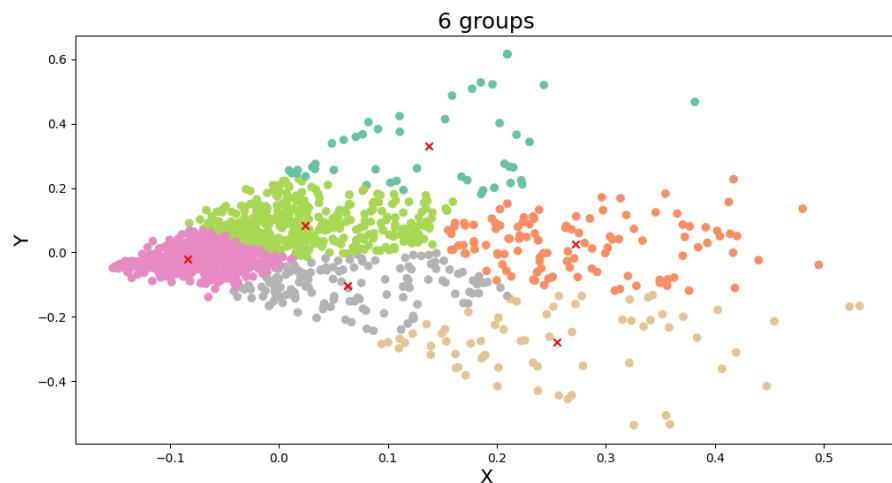
ภาพที่ 47 จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม



ภาพที่ 48 จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม



ภาพที่ 49 จัดกลุ่มข้อมูลจำนวน 7 กลุ่ม



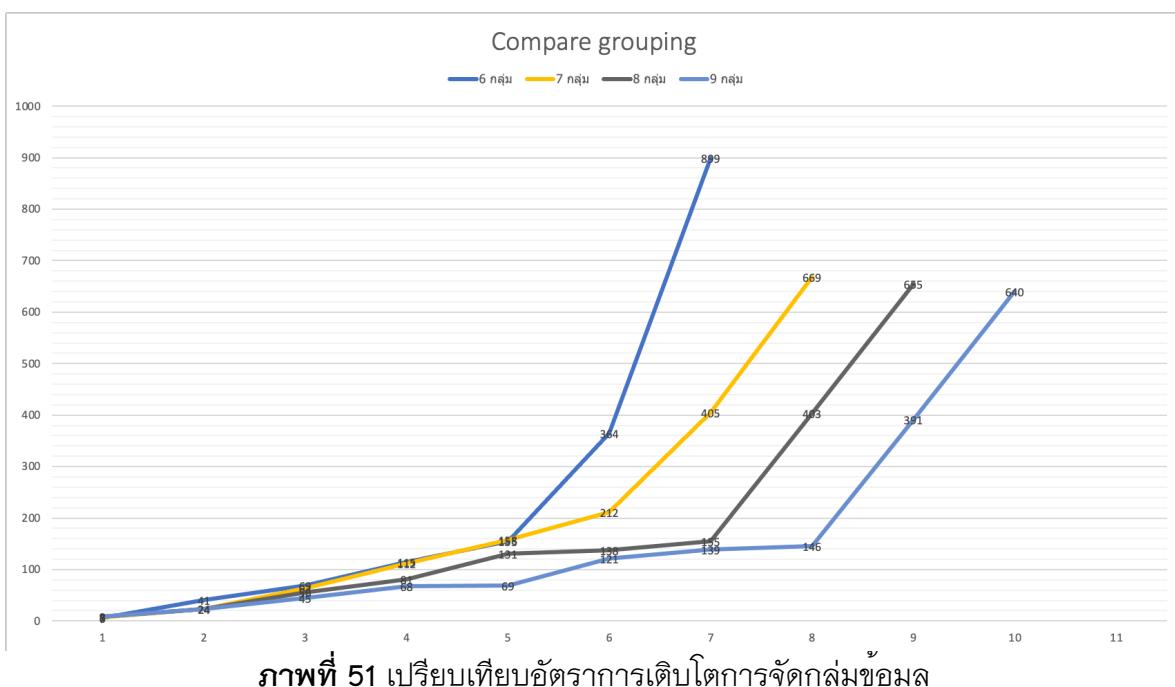
ภาพที่ 50 จัดกลุ่มข้อมูลจำนวน 6 กลุ่ม

ทำการทดลองนับจำนวนรายการบริษัทแต่ละประเภทในแต่ละกลุ่มเพื่อค่าเฉลี่ยและการกระจายตัวของข้อมูลบริษัท

ตาราง 15 แสดงการนับจำนวนบริษัทแต่ละประเภทในการจัดกลุ่มจากทั้งหมด 1,643 รายการ

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

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



จากภาพที่ 47-50 จะเห็นได้ว่าการกระจายตัวของข้อมูลนั้นมีความแตกต่างกันสังเกตได้จากสีที่ระบุตำแหน่งของกลุ่มในแต่ละภาพโดยพำน代เมื่อเปรียบเทียบการแบ่งกลุ่มที่ 6 และ 8 จะเห็นว่าขอบเขตของข้อมูลของการแบ่ง 6 กลุ่มนั้นมีความกว้างมากและห่างไกลจากจุดกึ่งกลางของข้อมูล (Centroids point) จึงประมานได้ว่าการแบ่งกลุ่มที่จำนวน 6 กลุ่มนั้นอาจไม่ได้ประสิทธิภาพความแม่นยำมากพo และในตรงกันข้ามการแบ่งกลุ่มข้อมูลที่ 9 กลุ่มนั้นจะเห็นได้ว่าขอบเขตของข้อมูลนั้นเล็กมากจนถึงทับซ้อนกันในแต่ละกลุ่มถึงแม้ขอบเขตของข้อมูลจะอยู่ใกล้ๆกันก็ตาม ดังนั้นผู้วิจัยจึงตัดตัวเลือกการแบ่งกลุ่มข้อมูลที่ 6 กลุ่มออกเหลือเพียง 7 8 และ 9 กลุ่ม

และจากภาพ 51 จะเห็นได้ว่าการเติบโตของข้อมูลเพิ่gn นำปริมาณบริษัทในแต่ละประเภทมาเรียงจากน้อยไปมากนั้น อัตราการเติบโตของการจัดกลุ่มข้อมูลที่ 7 กลุ่มนั้นมีอัตราการเติบโตที่คงที่มากที่สุด

จากนั้นทดลองทำการสุมเรียกข้อมูลในแต่ละกรณีอุปกรณ์เพื่อประกอบการตัดสินใจที่จะเลือกจำนวนกลุ่มของข้อมูลดังภาพด้านล่างที่ 47-50

จากการทดลองสุ่มเรียกข้อมูลหลาย ๆ ครั้งพบว่าจำนวนของการจัดกลุ่มที่มีค่าข้อมูลทับช้อนกันน้อยที่สุดอยู่ที่ 7 กลุ่มทำให้ผู้วิจัยเลือกที่จะแบ่งกลุ่มข้อมูลที่ 7 กลุ่มแต่ทั้งนี้ก็ยังมีข้อมูลที่ทับช้อนกันอยู่บ้างเล็กน้อยซึ่งอยู่ในระดับที่รับได้ และเมื่อทำการแบ่งกลุ่มข้อมูลเรียบร้อยแล้วก็สามารถบันทึกข้อมูลพร้อมกับ ID ของกลุ่มเพื่อนำไปสำหรับงานข้อมูลและใช้งานต่อไป

#### 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. งานปรีกษา/toที่
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@Tingrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh)                                361 ..final_project (-zsh)
[ 28% ] [ 56 GB ] [ 20:24:16 ]                         [ 20:24:16 ]
~/Desktop/final_project develop 12 ?                | ~/Desktop/final_project
python cluster.naming.py
Group: 0
1: บริษัท Platform ชั้นนำและนวัตกรรม เป้าหมายการเข้าสู่ห้องน้ำในประเทศไทยและอาเซียนในปีหน้า
2: บริษัทค้าปลีกอิเล็กทรอนิกส์ (Software-as-a-Service) Hospitality Industry Management Systems (Front Office, Point of Sale, Call Accounting, Back Office with Accounts Payable, Accounts Receivable, General Ledger, and Inventory Control)
3: บริษัทผู้ผลิตและจัดจำหน่ายสินค้าสำหรับบ้านและสวน เช่น กระถางต้นไม้ ห้องน้ำ
4: บริษัทผลิตกระดาษless (Paperless Solution) และเครื่องพิมพ์ DForm Paperless System ให้เช่น Cloud based (คลาวด์การพิมพ์ข้อมูลเอกสารดิจิทัล) และ On-Premise (พิมพ์การพิมพ์แบบดั้งเดิม)
5: บริษัท Software as a service on Cloud แพลตฟอร์มจัดการธุรกิจของธุรกิจขนาดย่อม
6: บริษัทออกแบบและพัฒนาระบบเทคโนโลยีทางด้าน VR ที่ชื่อ VIRTUAL REALITY (VR) VR Sphere ความท้าทายที่มากที่สุด VR, Software, Hardware, Mechanic และ Product Design, รวม Smart Workplace Solution (Meet in Touch, CO Desk, Visitor) (Visitor Management System, Hotel Desk System, Meeting Room Reservation System, Locker, On-Premise Face Recognition System)
7: บริษัทค้าปลีก Software as a Service ที่เน้นก้าวไปสู่การขายอิเล็กทรอนิกส์ Big Data ที่เก็บรวบรวมข้อมูลเพื่อให้คำแนะนำ (Social Monitoring, Social Engagement, Web-based social intelligence solution, integrating social listening, monitoring, engagement and customer service)
8: บริษัทค้าปลีกอิเล็กทรอนิกส์ E-Library (E-Library) พัฒนาและขยายห้องสมุดอิเล็กทรอนิกส์ (Auto-Lib) และพัฒนาศูนย์เรียนรู้ได้รับเงิน และสร้างห้องเรียน Cloud รวมถึง Digital Knowledge Center, Learning Management System และ Library Management System
9: บริษัทค้าปลีกและห้องอาหารที่มีชื่อเสียง เช่น บีบีพาร์ค ให้บริการโซลูชันด้าน Digital Infrastructure โดยมีพันธมิตร :
Ascend Money : Fintech & E-Payments : TrueMoney, Ascend Nano
Ascend Commerce : E-Commerce / E-Procurement / Fulfillment : WeLoveShopping, WeMail, WeFresh, GoodChoiz.com, Pantanavij, Ascend Travel, Egg Digital, Aden Digital Enablers : Data Center & Cloud Service Provider : TrueIDC
10: บริษัทซอฟต์แวร์ที่ควบคุมอุปกรณ์ ในโทรศัพท์มือถือและ F10 และ RapidSTM32 Blockset หรือแม่บอร์ดผู้ติดต่อ (Custom Blockset Development), Smart Sensor

~/Desktop/final_project develop 12 ?                [ 20:24:16 ]
```

ภาพที่ 52 ตัวอย่างข้อมูลในกลุ่ม 0

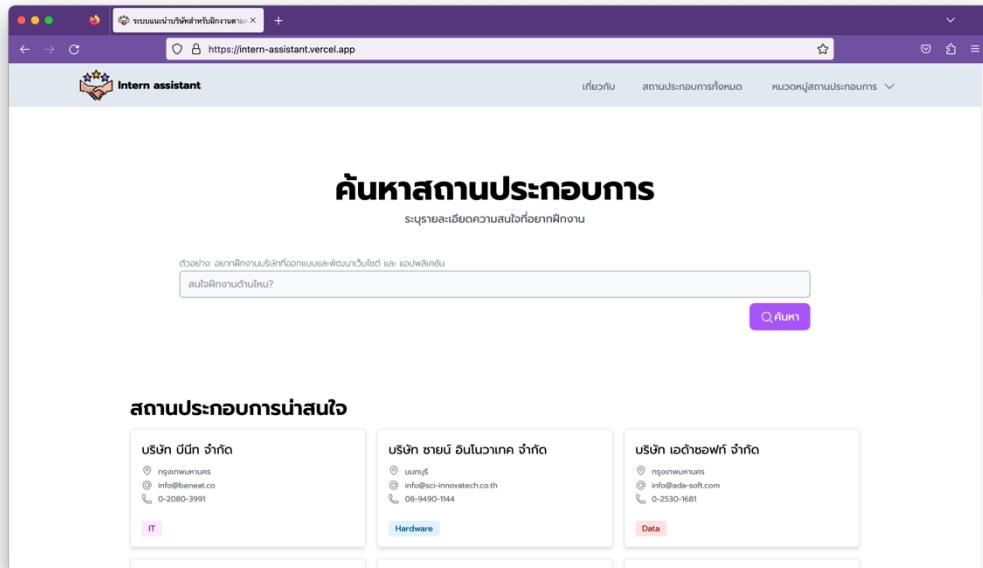
### ภาพที่ 53 ตัวอย่างข้อมูลในกลุ่ม 2

จากการทำกิจกรรมเรียนรู้ชื่อมูลดูทั้ง 7 กลุ่ม รายชื่อประเภทงานที่สามารถใช้ตั้งชื่อกลุ่มชื่อมูลโดย  
อ้างอิงของเว็บไซต์ [th.jobsdb.com](http://th.jobsdb.com) มีดังนี้

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

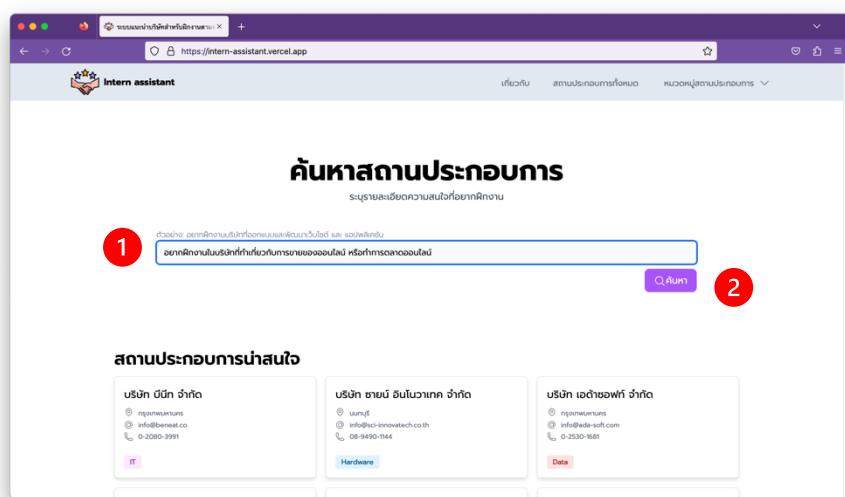
## 4.2 ขั้นตอนการใช้งานสำหรับผู้ใช้งาน

### 4.2.1 หน้าแรกเว็บไซต์ Intern-assistant



ภาพที่ 54 หน้าแรกเว็บไซต์ Intern–assistant

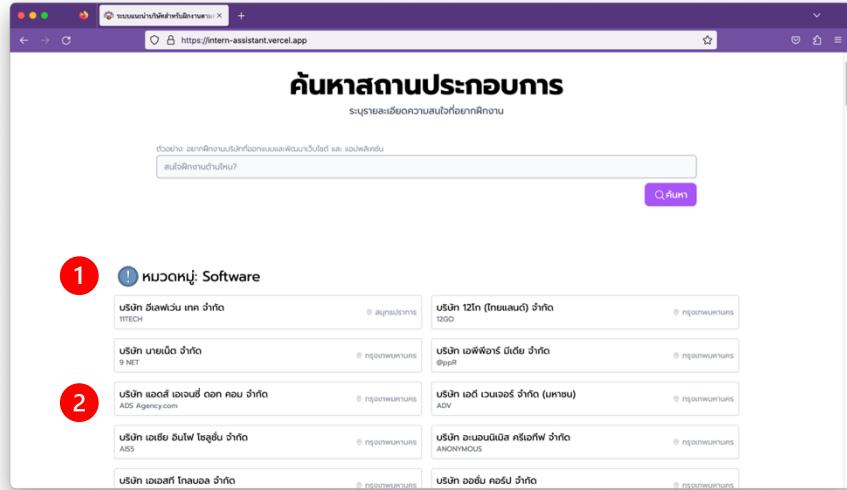
### 4.2.2 คนหาบริษัท



ภาพที่ 55 คนหาบริษัท

หมายเหตุ 1 พิมพ์ค้นหาบริษัทที่ต้องการฝึกงานด้วยรายละเอียดความสนใจรูปแบบใดก็ได้  
หมายเหตุ 2 กดปุ่มค้นหา

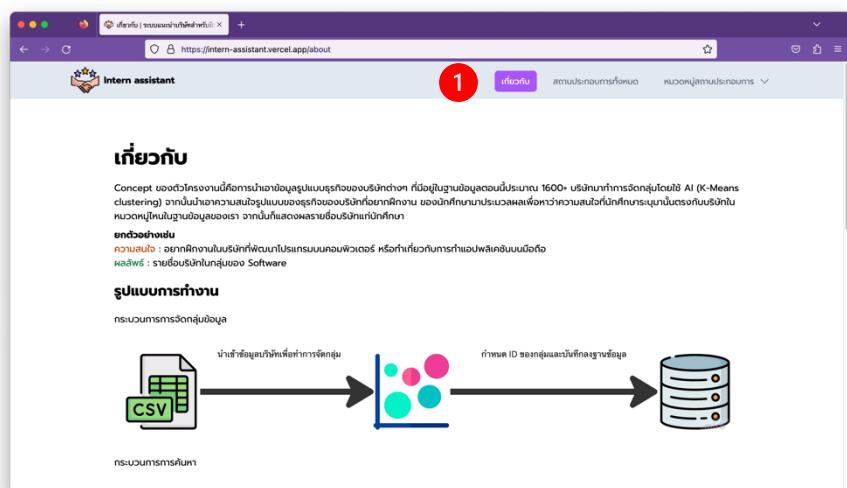
#### 4.2.3 หน้าแสดงผลลัพธ์การค้นหา



## ภาพที่ 56 หน้าแสดงผลลัพธ์การค้นหา

หมายเลขอ 1 ชื่อของกลุ่มข้อมูล  
หมายเลขอ 2 รายชื่อปริยัทท์หมดที่อยู่ในกลุ่ม

#### 4.2.4 หน้าเกี่ยวกับ



มาตรา 57 แห่งรัฐธรรมนูญ

หมายเลขอ 1 เมนหน้าเกี่ยวกับ แสดงข้อมูลเกี่ยวกับระบบและการทำงาน

#### 4.2.5 หน้าแสดงรายชื่อบริษัททั้งหมด

The screenshot shows a web browser displaying the Intern Assistant application. The URL is https://intern-assistant.vercel.app/company/all. At the top, there is a search bar with the placeholder 'ค้นหาบริษัท'. Below the search bar, the page title is 'สถานประกอบการทั้งหมด'. The main content area displays a grid of company cards. Each card contains the company name, logo, and location. A red circle labeled '1' points to the search bar, and another red circle labeled '2' points to the list of companies.

บริษัท ลูกข่างอุ่น จำกัด	บริษัท 10 กันยา จำกัด 10 September Co., Ltd.	บริษัท วันเมือง ไทยเน็ต จำกัด 101 Global Co., Ltd.
บริษัท 11-softfotech จำกัด 11-softfotech System Co., Ltd.	บริษัท ดีลฟอร์เวน เทค จำกัด Eleven Tech Co., Ltd.	บริษัท 12GO (ไทยแลนด์) จำกัด 12GO (Thailand) Co., Ltd.
บริษัท วนกูโกลด์ บันเดอร์กรุ๊ป จำกัด Onewgold Intergroup Co., Ltd.	สำนักงานบันได 147 แห่ง สุรินทร์ 147 KP Studio Ltd., Part	บริษัท วันดี จำกัด One DD Hub Co., Ltd.
บริษัท วัน เมดิคอลพาร์ท จำกัด 1 Medisoft Expert Co., Ltd.	บริษัท วันไธล็อก ไซส์บิ๊กแอดดิชัน จำกัด 1 Hotel Solutions.com Co., Ltd.	บริษัท ภูริญช์ (ประเทศไทย) จำกัด 2C2P (Thailand) Co., Ltd.
บริษัท กฎ โซลูชัน แอนด์ เซอร์วิส จำกัด 2J Solution & Service Co., Ltd.	บริษัท 360 บันบีฟ จำกัด 360 Innovative Co., Ltd.	บริษัท นิรવิจัย จำกัด 3View Group Co., Ltd.
บริษัท กฟิว บันไฟฟ้า (ประเทศไทย) จำกัด	บริษัท ไฟร์ ศิลป์แล็บ ถีรย์ จำกัด FIRE Silp Lab Co., Ltd.	บริษัท พลังแมก จำกัด

ภาพที่ 58 หน้าแสดงรายชื่อบริษัททั้งหมด

หมายเหตุ 1 เมนูหน้ารายชื่อบริษัททั้งหมด

หมายเหตุ 2 รายชื่อบริษัททั้งหมด

#### 4.2.6 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

The screenshot shows a web browser displaying the Intern Assistant application. The URL is https://intern-assistant.vercel.app/company/category/0. At the top, there is a search bar with the placeholder 'ค้นหาบริษัท'. Below the search bar, the page title is 'หมวดหมู่: Data'. The main content area displays a grid of company cards. Each card contains the company name, logo, and location. A red circle labeled '1' points to the search bar, another red circle labeled '2' points to the category dropdown menu, and a third red circle labeled '3' points to the list of companies.

บริษัท 11-softfotech จำกัด 11-softfotech SYSTEM	บริษัท ภูริญช์ (ประเทศไทย) จำกัด 2C2P
บริษัท กฟิว บันไฟฟ้า (ประเทศไทย) จำกัด 2J Infotech	บริษัท นายดี จำกัด 9 Net
บริษัท ดีโซล คอมพิวเตอร์ (ประเทศไทย) จำกัด DOS Computer	บริษัท เอ-ไฮส์ จำกัด A-Host
บริษัท เมโด อิวานิช แอนด์ แอนดี้พีเพิลเซอร์ช จำกัด AI Group	บริษัท แอดวานซ์ บีซีพี ไอส์บีซี คอมปьюเตอร์ จำกัด ASBC
บริษัท เมดิคอลพาร์ท จำกัด ACETERA	บริษัท อาดาстра (ไทยแลนด์) จำกัด ADASTRA
บริษัท แอดวานซ์ ดิจิตอล บลสแลด จำกัด ADB	บริษัท เมดิคอลพาร์ท จำกัด ADEPTUS
บริษัทเมดิคอล คลินิกบันกอกสิน จำกัด AERIAL	บริษัท แอร์พอร์ตส์ จำกัด AIRPORTELS

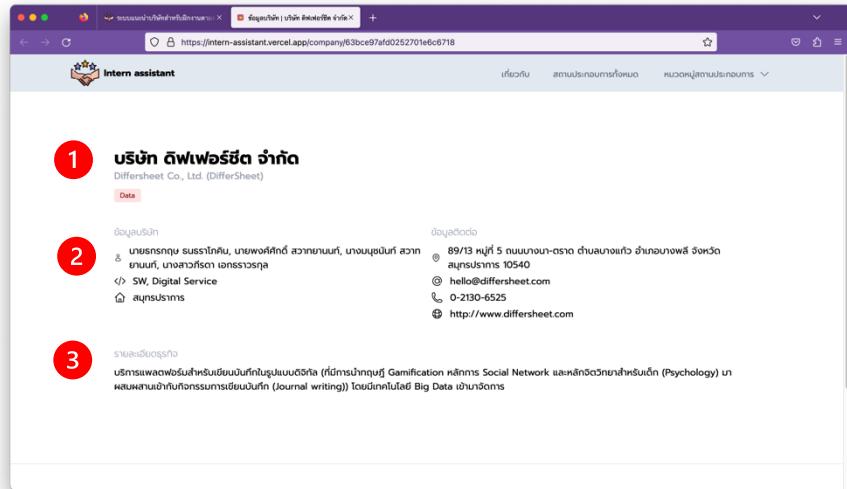
ภาพที่ 59 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หมายเหตุ 1 เมนูรายชื่อกลุ่มบริษัท

หมายเหตุ 2 รายชื่อกลุ่มบริษัท

หมายเหตุ 3 รายชื่อบริษัทในกลุ่ม

#### 4.2.7 หน้ารายละเอียดบริษัท



ภาพที่ 60 หน้ารายละเอียดบริษัท

#### 4.3 การวัดค่าความคล้ายคลึง

Cosine similarity เป็นเทคนิคที่นำมาใช้หาความคล้ายคลึงระหว่างความสนใจของผู้ใช้และข้อมูลบริษัทที่อยู่ในฐานข้อมูลยิ่งค่า Cosine similarity เข้าใกล้ 1 แสดงว่าประโยชน์มีความคล้ายคลึงกับข้อมูลบริษัทในกลุ่มนั้นมากดังภาพที่ 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: ออกแบบเว็บไซต์ด้วย react js ทำเกี่ยวกับการเขียนเว็บ การตลาดออนไลน์ด้วย และ SEO
cluster: 6
cosine similarity: 0.07871941760441518
~/desktop/final_project develop !3 ?2
```

ภาพที่ 61 ตัวอย่างการคำนวณค่า Cosine similarity

##### 4.3.1 การหากลุ่มที่มีความคล้ายคลึงมากที่สุด

1. คำนวณค่า Cosine similarity ทุกกลุ่มจากความสนใจของผู้ใช้ผ่าน API ที่สร้างไว้เพื่อคำนวณโดยเฉพาะ

The screenshot shows a Postman workspace titled "My Workspace". A collection named "main testing" contains a POST request to "test cluster api". The request URL is <https://lamonze.tech/search>. The "Body" tab shows a JSON payload:

```

1 {
2   "keyword": "ອາກົດໃຈການໜີ້ຫຼືກົດຄອບນັບແລະພົມໄວໃຈສີ ແລະ ແນວຍເກີນ"
3 }

```

The response status is 200 OK, with a response body containing:

```

1 {
2   "cosine_similarity": [
3     0.03592689333934359,
4     0.0,
5     0.16082658037980805,
6     0.20863117753192784,
7     0.204765186452510848,
8     0.3397083961645713,
9     0.04001828248279789
10   ],
11   "max_cosine_similarity": 0.3397083961645713,
12   "cluster": 5
13 }

```

ภาพที่ 62 ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าความคล้ายคลึง

The screenshot shows a Postman workspace titled "My Workspace". A collection named "Calculate cosine similarity" contains a POST request to "searchcompany". The request URL is <https://lamonze.tech/searchcompany>. The "Body" tab shows a JSON payload:

```

1 {
2   "keyword": "ອາກົດໃຈການໜີ້ຫຼືກົດຄອບນັບແລະພົມໄວໃຈສີ ແລະ ແນວຍເກີນ ແລະ ຂໍເປົ້າເປົ້າການຮຽນຂອງໂຄສະນາ"
3 }

```

The response status is 200 OK, with a response body containing:

```

1 {
2   "short_company": "MANINNOVATION",
3   "th_company_name": "ມັນ ໄວ ນິວອຸວັນ ຈຳເກີນ",
4   "eng_company_name": "Man Innovation Co., Ltd.",
5   "type_business": "SW, Digital Content",
6   "detail": "ຫຼັບຜົນດີ່ນເວັບໄວ້ (Web Application), ລວມທີ່ສັນຕິພິບ (Logo Design), ລວມທີ່ສັນຕິພິບ (Graphic & Stationery Design)", 
7   "owner": "ວຽກທີ່ສັນຕິພິບ ໃຫຍວງ ດີເລີ່ມຕົວໄກ້, ນາງສູງລັດ ດີເລີ່ມຕົວໄກ້",
8   "province_base": "ຖານວຽກ",
9   "address": "30 ຂອບພະຍາຍື 16 ເມນະວຽກ ເມນະວຽກ ກູດເມນະວຽກ 10520",
10  "phone_number": "0-2727-7274",
11  "email": "thamee@maninnovation.com",
12  "website": "http://thumweb.maninnovation.com",
13  "cluster": "5",
14  "company_id": "e11d709b-bb0f-11ed-b954-18310f2091eb",
15  "cluster_name": "Network"
16  "short_company": "Deeboon",
17
18
19

```

ภาพที่ 63 ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่คล้ายที่สุด

## บทที่ 5

### สรุปผลการดำเนินงาน

การพัฒนาระบบแนะนำบริษัทสำหรับผู้ใช้งานตามความสนใจ ด้วยเทคโนโลยีปัญญาประดิษฐ์ ในครั้งนี้สามารถสรุปการดำเนินงาน ปัญหาที่เกิดขึ้นระหว่างการดำเนินงาน ขอเสนอและแนวทางพัฒนาต่อไปดังนี้

#### 5.1 สรุปผลการดำเนินงาน

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

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 ข้อมูลใหม่ที่จะเข้ามารวมกับข้อมูลเดิมที่มีอยู่แล้วต้องนั่งข้อมูลของแต่ละบริษัททีละๆ กันเปลี่ยนกลุ่มไปทุกครั้งทั้งมีการเพิ่มข้อมูลใหม่

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

5.2.3 ข้อมูลที่ได้รับมาเมื่อมีคำที่สอดคล้องไม่สามารถแก้ไขได้ด้วยผู้พัฒนาเอง การทำ Clustering นั้นอาจจะไม่ได้แยกแยะข้อมูลได้ดียกตัวอย่างเช่น ประโยชน์ที่มีคำว่า “แอปพลิเคชัน” อาจจะอยู่ในกลุ่มกับประโยชน์ที่มีคำว่า “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 พัฒนาให้รองรับการกรองตัวเลือกที่จะคนหา
- 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/คุณมีอ/73-คืออะไร/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-คืออะไร>

(2022, 3 12). Retrieved from Use Case Diagram: <https://www.mindphp.com/บทความ/31-ความรู้ทั่วไป/6870-use-case-diagram.html>

การวิเคราะห์ระบบและการออกแบบ System Analysis and Design (ซิลเด็ม อนาคต แอน ดิโอล์น). (2022, 3 30). Retrieved from mindphp: <https://www.mindphp.com/บทความ/31-ความรู้ทั่วไป/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 จะเข้ามาช่วยองค์กรของคุณได้อย่างไร? (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). เทคนิคการคัดเลือกกลุ่มโหลดรายอาทิตย์สำหรับ รองรับแผนการติดตั้งระบบผลิตไฟฟ้าพลังงานแสงอาทิตย์บนหลังคาเพื่อเพิ่มคาดารชนี ประสิทธิภาพการใช้พลังงานไฟฟ้า Load Clustering Technique Application to PV Solar Rooftop Installation Planning for Improving Energy Efficiency, 134–148.

ปราณีย์ พึงวิชา, อ. ท. (2019). ศึกษาการแบ่งกลุ่มพฤติกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่าน เครือข่ายสังคมออนไลน์ Clustering of Jewellery Purchasing Behaviour through Social Network, 213–224.

จักรินทร์ สันติรัตนภักดี, ศ. น. (2021). การออกแบบและพัฒนากระบวนการจำแนกข้อร้องเรียนรถ โดยสารสาธารณะเพื่อติดแท็กปัญหาการให้บริการ, 77–90.

วุฒิชัย, ว. (2013). การเปรียบเทียบวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างการเขียนกับ โครงสร้างพยัญชนะ 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-อักษรจะกิน-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-ได-แยกประเภทอีเมลล์ด้วยพลังฟอร์ซ-66b8bdff2e42>

រាជធានីភ្នំពេញ

ក្រសួងពេទ្យ

ជាតិ

ກາດພນວກ ກ.  
ຄມືອກາຣຕິດຕັ້ງ

การเตรียมข้อมูลและติดตั้งโปรแกรมสำหรับการพัฒนาระบบแนะนำบริษัทสำหรับฝึกงานตามความสนใจ ด้วยเทคโนโลยีปัญญาประดิษฐ์

## 1. การจัดกลุ่มข้อมูลด้วยเทคนิคทางปัญญาประดิษฐ์เคลื่อนที่ (K-Means)

1.1 ทำการดาวน์โหลด Repository จาก [https://github.com/slapexs/final\\_project](https://github.com/slapexs/final_project) ด้วยการใช้คำสั่ง `git clone https://github.com/slapexs/final_project` เพื่อดownloadไฟล์มาไว้ในเครื่อง

```
44% 5.5 GB ~1 .project/report (~zsh) ⚡ 82 +  
~/Desktop  
~/desktop ➜ git clone https://github.com/slapexs/final_project.git  
Cloning into 'final_project'...  
remote: Enumerating objects: 821, done.  
remote: Counting objects: 100% (117/117), done.  
remote: Compressing objects: 100% (63/63), done.  
remote: Total 821 (delta 59), reused 110 (delta 53), pack-reused 704  
Receiving objects: 100% (821/821), 51.74 MiB | 6.05 MiB/s, done.  
Resolving deltas: 100% (401/401), done.  
12s ⚡ 16:42:53
```

ภาพที่ 64 ดาวน์โหลด Repository ด้วยคำสั่ง git clone

1.2 บันทึกไฟล์ข้อมูลบริษัทนามสกุล .csv ในโฟลเดอร์ data\_csv



ภาพที่ 65 แสดงการบันทึกไฟล์ข้อมูลบริษัทในโฟลเดอร์

### 1.3 เปิดไฟล์ clustering.py ในโปรแกรม Text editor

ภาพที่ 66 โค้ดคำสั่งในไฟล์ clustering.py ใช้ในการจัดกลุ่มข้อมูล

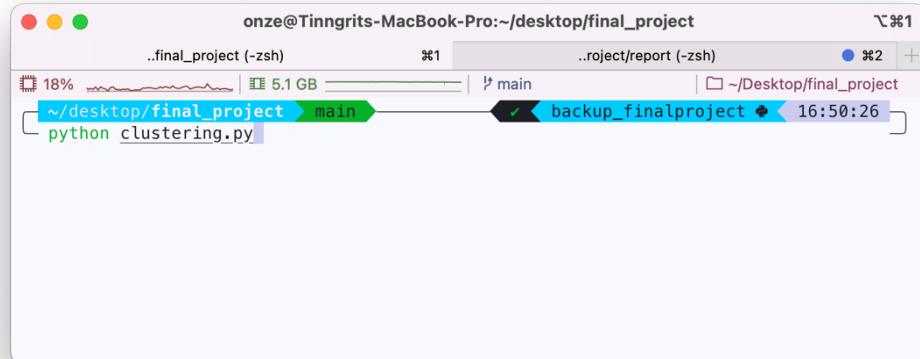
1.4 ทำการกำหนดจำนวนของกลุ่มข้อมูลที่ต้องการลงในตัวแปร k

```
● ● ●  
1 k = 7  
2 kmeans = KMeans(n_clusters=k, random_state=1)  
3 # Fit model  
4 kmeans.fit(df_tfidf[['x_value', 'y_value']])  
5 clusters = kmeans.labels_
```

ภาพที่ 67 แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล

จากการที่ 62 แสดงการจัดกลุ่มข้อมูลด้วยเคมีนท์ (K-Means) โดยการที่กำหนดจำนวนกลุ่มข้อมูลที่ต้องการลงในตัวแปร k จากนั้นทำการจัดกลุ่มข้อมูลด้วยฟังก์ชัน KMeans และคืนค่ากลับมา yang ตัวแปร kmeans และเก็บป้ายชื่อของกลุ่มลงในตัวแปร clusters

## 1.5 เรียกใช้ไฟล์เพื่อทำการจัดกลุ่มข้อมูล



ภาพที่ 68 แสดงการรันคำสั่งจัดกลุ่มข้อมูลใน Terminal

## 1.6 ดูไฟล์ clustered\_company.csv ในโฟลเดอร์ document ที่เป็นผลลัพธ์การจัดกลุ่มข้อมูล

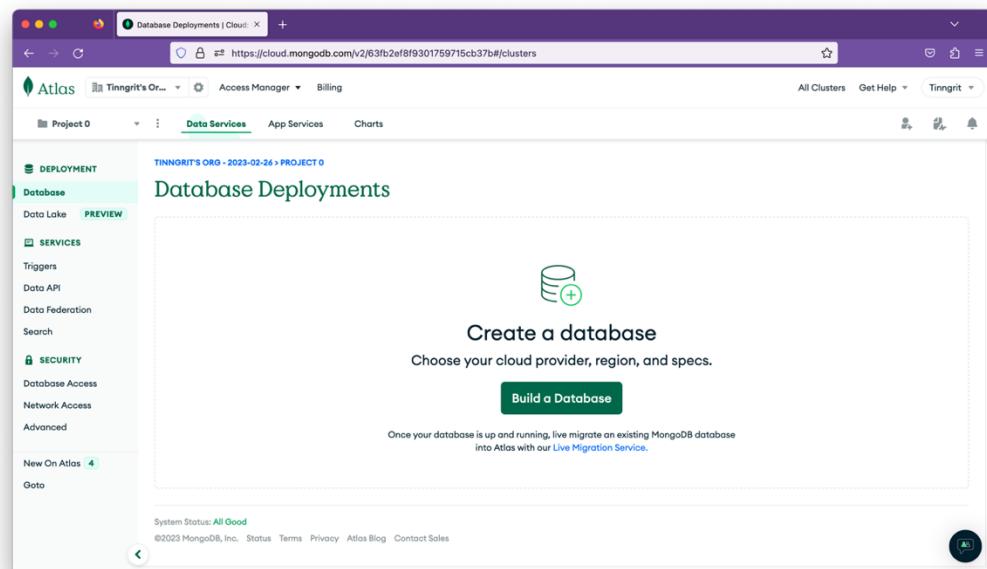
	h_company	Eng_compan	Type_busine	Type_techno	Product	Type_innovat	Detail	Owner	Province_bas	Address	Phone_num	Email	unnamed						
1	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นางสาว กัญญาดา กระเทียมหนานดา	18,20 ซอยกรุงเทพฯ	09-0004-000	support@ooc.co.th										
2	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย อรุณรัตน์ นิตย์บูรณ์	กรุงเทพฯ	1168 ถนนพระราม 2-2026-3238	contact@drive										
3	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย พัฒนา โชค	กรุงเทพฯ	พืช 201 อาคาร 1	0-2120-6013	customer@venue									
4	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
5	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
6	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
7	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
8	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
9	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
10	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
11	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
12	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
13	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
14	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
15	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
16	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
17	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
18	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
19	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									
20	ห้องทำงาน Drivemate Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	แมลติมีเดีย	บริการ Platform	นาย ภูมิธรรม วิชัย	กรุงเทพฯ	พืช 201 อาคาร 1	0-2080-3966	customers@hi									

ภาพที่ 69 แสดงไฟล์ clustered\_company.csv

## 2. การสร้างคลัสเตอร์ (Cluster) ของฐานข้อมูล mongoDB บนเว็บไซต์

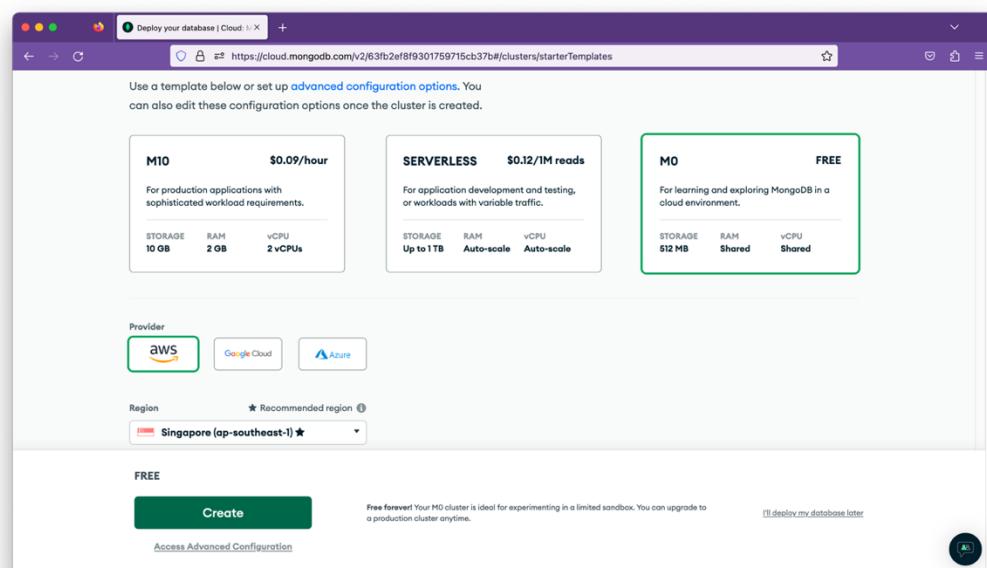
2.1 การใช้งาน Cloud MongoDB เข้าเว็บไซต์ <https://www.mongodb.com/> จากนั้นเข้าสู่ระบบ

2.2 สร้าง Cluster โดยการกดปุ่ม Build a Database



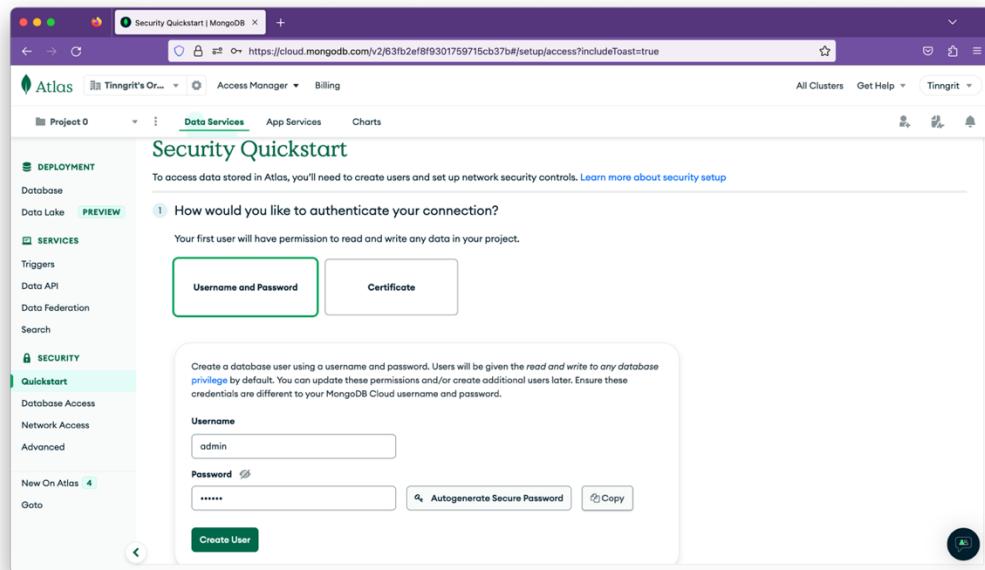
ภาพที่ 70 แสดงหน้าการจัดการ Cluster MongoDB

2.3 เลือกการตั้งค่าของ Cluster และตั้งชื่อ จากนั้นกดปุ่ม Create เพื่อสร้างฐานข้อมูลโดยที่เลือกผู้ให้บริการเป็น AWS และเลือกภูมิภาคเป็นประเทศไทย



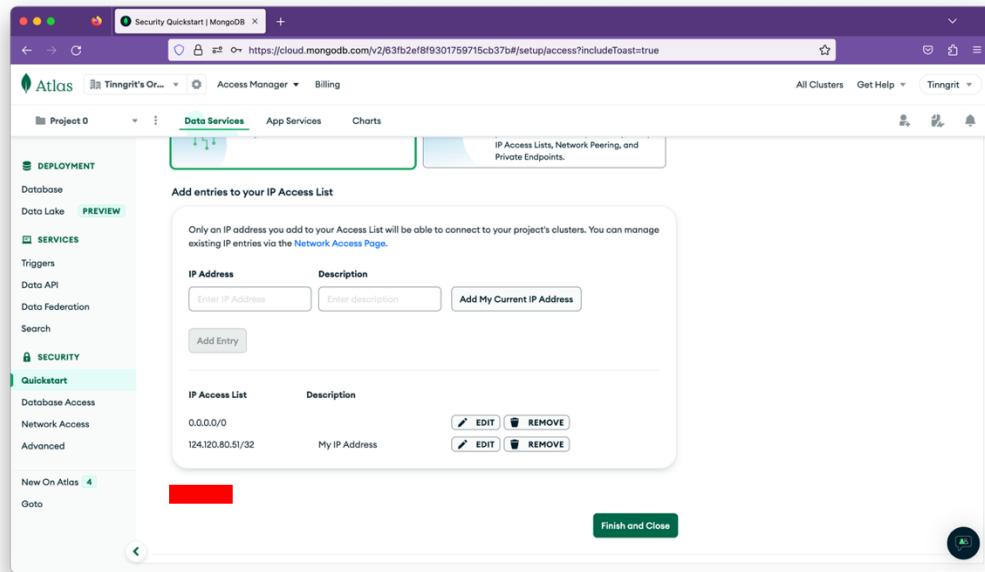
ภาพที่ 71 แสดงหน้าตั้งค่าและสร้าง Cluster MongoDB

## 2.4 สร้างบัญชีสำหรับใช้งานฐานข้อมูลตั้งค่า username และ password จากนั้นกดปุ่ม Create User



ภาพที่ 72 แสดงหน้าสร้างบัญชีสำหรับจัดการฐานข้อมูล

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



ภาพที่ 73 แสดงหน้าเพิ่ม IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้

### 3. การสร้างฐานข้อมูลในโปรแกรม mongdb (MongoDB)

#### 3.1 สร้างฐานข้อมูลและ Collection กดที่ชื่อ Cluster

The screenshot shows the MongoDB Atlas interface. On the left, there's a sidebar with options like Project 0, Data Services (selected), App Services, and Charts. Under Data Services, there are sections for Deployment (Database, Data Lake, PREVIEW), Services (Triggers, Data API, Data Federation, Search, SECURITY, Quickstart, Database Access, Network Access, Advanced), and New On Atlas. The main area is titled 'Database Deployments' and shows a single cluster named 'Cluster1'. It displays metrics such as R: 0, W: 0, Connections: 0, In: 0.0 B/s, Out: 0.0 B/s, and Data Size: 0.0 B. Below the metrics, it shows details like VERSION: 5.0.14, REGION: AWS / Singapore (ap-southeast-1), CLUSTER TIER: MD Sandbox (General), TYPE: Replica Set - 3 nodes, BACKUPS: Inactive, LINKED APP SERVICES: None Linked, and ATLAS SEARCH: Create Index. At the bottom, it says 'System Status: All Good'.

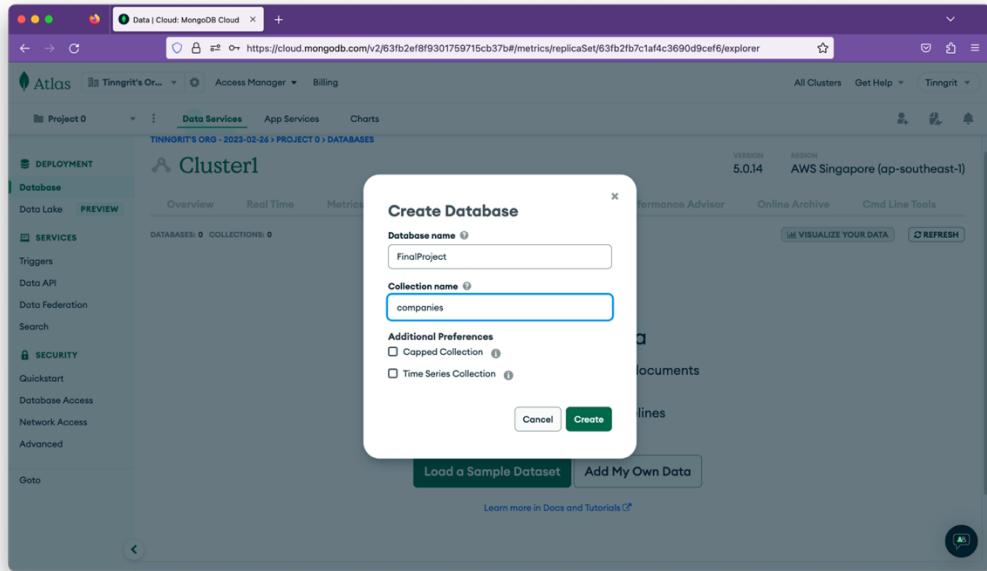
ภาพที่ 74 แสดงหน้าจัดการ Cluster MongoDB

#### 3.2 กดที่แท็บเมนู Collections และกดปุ่ม Add My Own Data เพื่อสร้างฐานข้อมูล

The screenshot shows the MongoDB Atlas interface for the 'Cluster1' cluster. The top navigation bar includes Project 0, Data Services (selected), App Services, and Charts. The left sidebar has the same structure as the previous screenshot. The main area is titled 'Cluster1' and shows 'OVERVIEW', 'REAL TIME', 'METRICS', and 'COLLECTIONS' tabs. The 'COLLECTIONS' tab is selected, showing 'DATABASES: 0' and 'COLLECTIONS: 0'. Below this is a section titled 'Explore Your Data' with a sub-section 'Find: run queries and interact with documents'. It also lists 'Indexes: build and manage indexes', 'Aggregation: test aggregation pipelines', and 'Search: build search indexes'. At the bottom of this section are two buttons: 'Load a Sample Dataset' and 'Add My Own Data'. A note at the bottom says 'Learn more in Docs and Tutorials'.

ภาพที่ 75 แสดงการสร้างฐานข้อมูล MongoDB

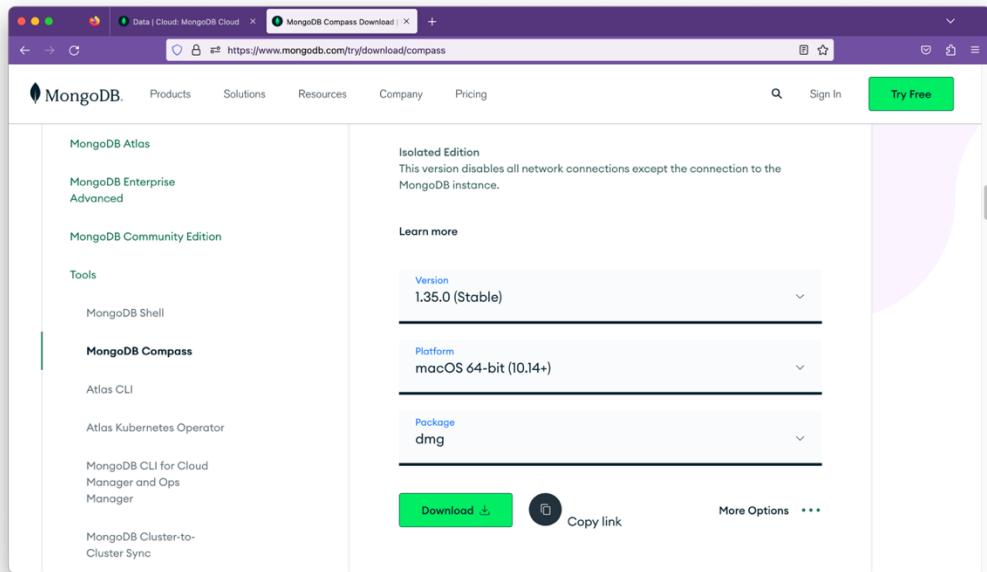
### 3.3 កំណត់ឱ្យរាជក្រឹតាគនាមូលនៃជីវិត Collection ដើម្បីនឹងរាជក្រឹតាគនាមូល



រាងទី 76 ផ្តល់ព័ត៌មានអំពីការសរុបរាជក្រឹតាគនាមូលនៃ Collection

### 3.3 គាន់និយោគពិន្ទុលិខិត្ត MongoDB compass ដើម្បីក្រុមហ៊ុន

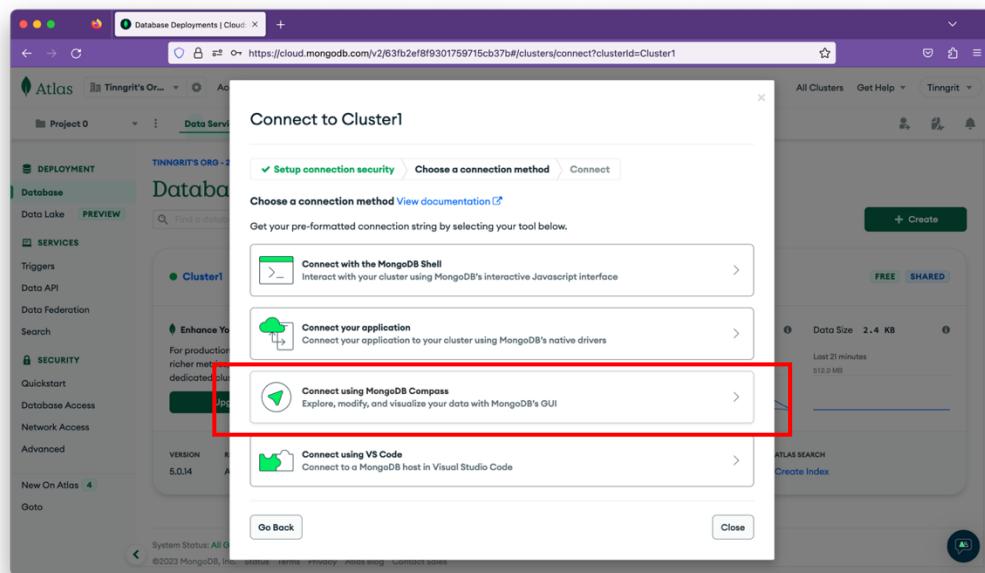
<https://www.mongodb.com/try/download/compass> ឡើង Version Platform និង Package តាមរបៀបប្រើប្រាស់។



រាងទី 77 ព័ត៌មានអំពីការក្រុមហ៊ុនគាន់និយោគពិន្ទុលិខិត្ត MongoDB compass

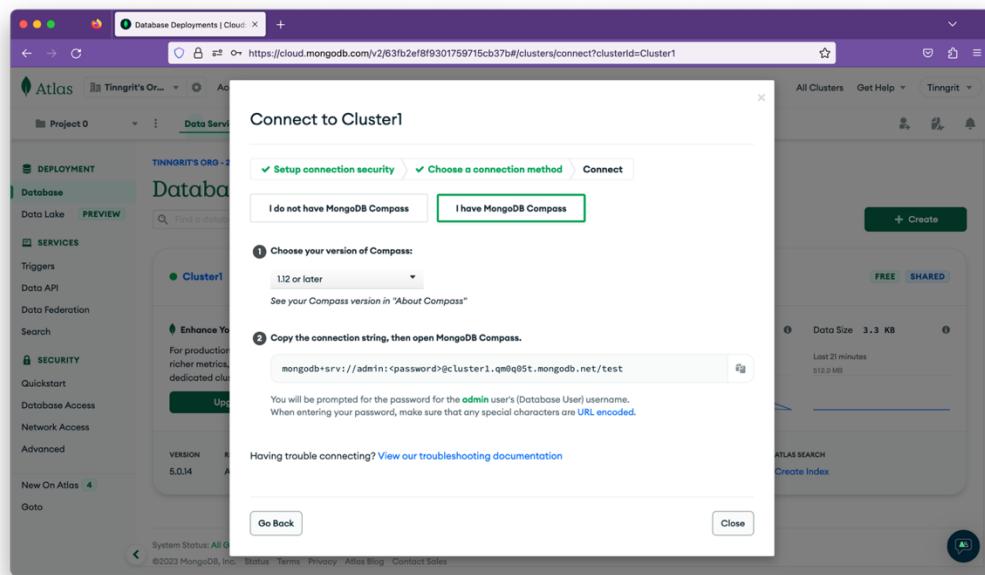
#### 4. การนำข้อมูลเข้าสู่ฐานข้อมูลของโกลเด้น (MongoDB)

4.1 เปิดเว็บไซต์หน้าจัดการ Cluster กดปุ่ม Connect และกดปุ่ม Connect using MongoDB Compass



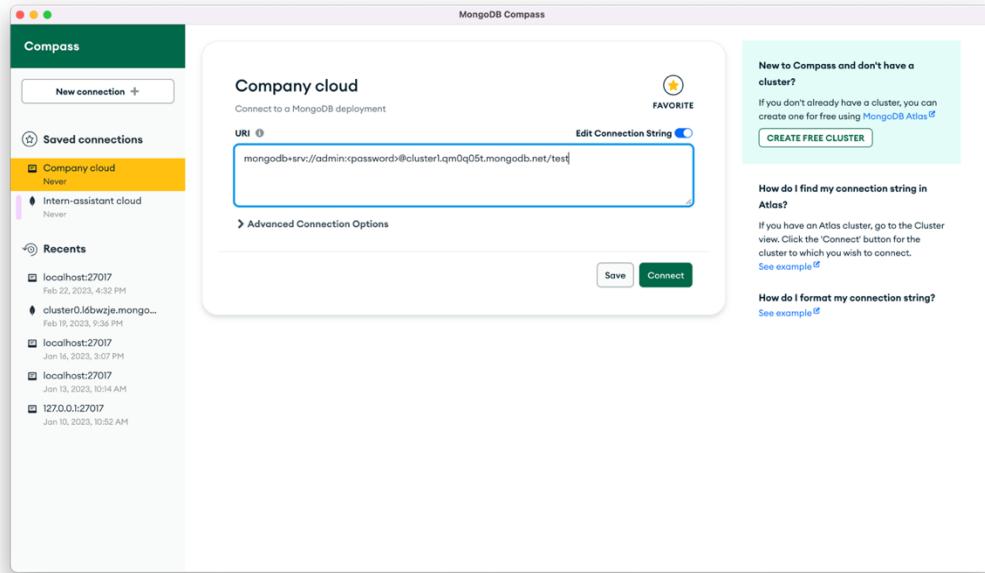
ภาพที่ 78 หน้าต่างเลือกเชื่อมต่อ กับ Cluster

4.2 กดปุ่ม I have MongoDB Compass และคัดลอก Connection string ไปข้อ 2



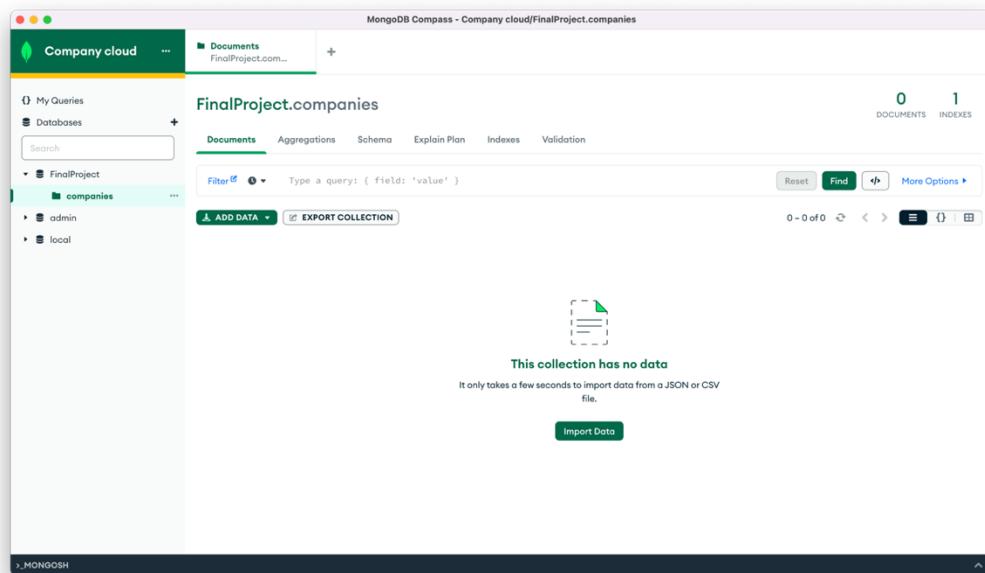
ภาพที่ 79 หน้าต่างข้อมูลการเชื่อมต่อ Cluster กับ MongoDB compass

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



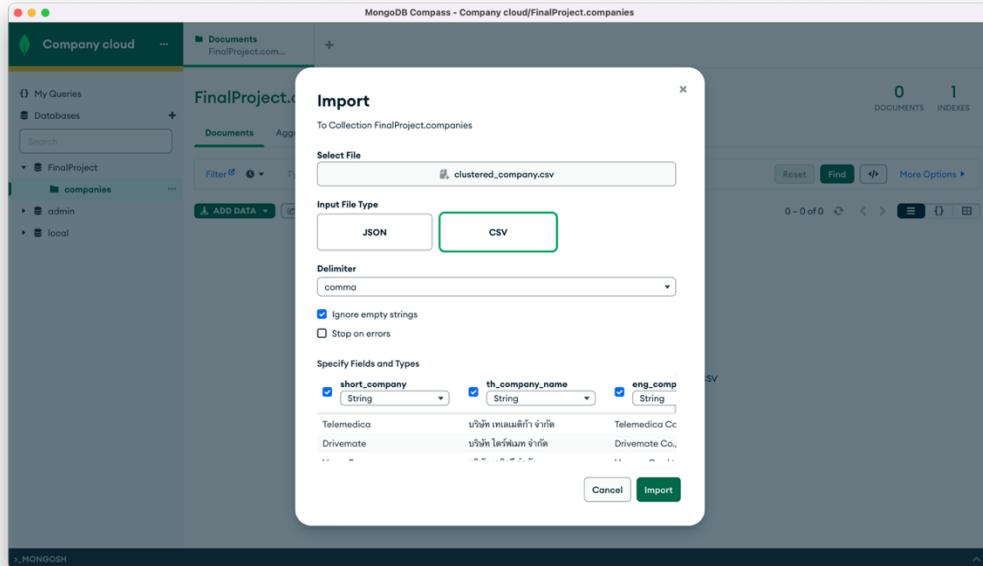
ภาพที่ 80 หน้าต่างโปรแกรม MongoDB compass สำหรับเชื่อมต่อ Cluster

4.4 เลือกฐานข้อมูลและ Collection ที่เมนูด้านซ้ายและกดปุ่ม Import Data



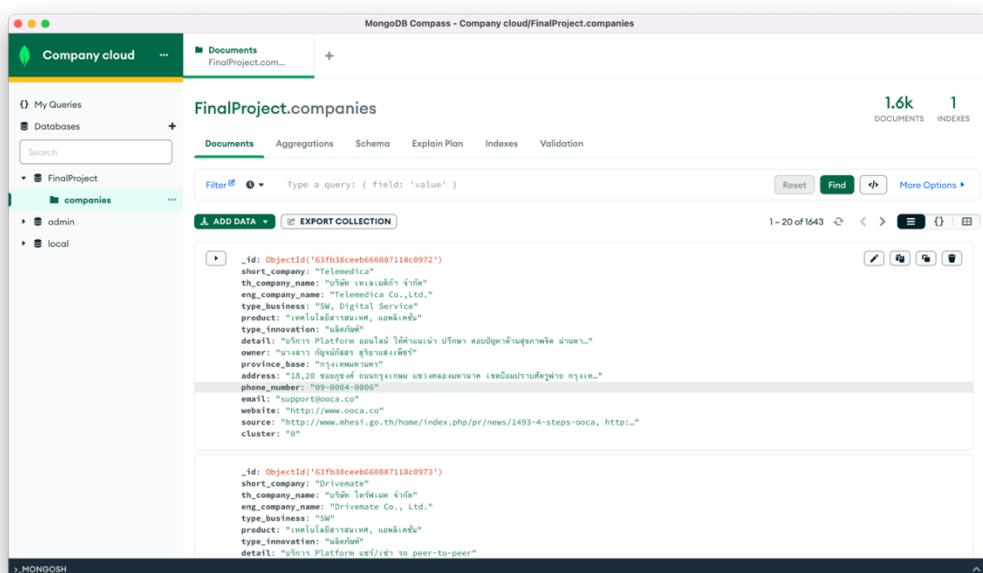
ภาพที่ 81 หน้าต่างโปรแกรมแสดงข้อมูลใน Collection

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



ภาพที่ 82 หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv

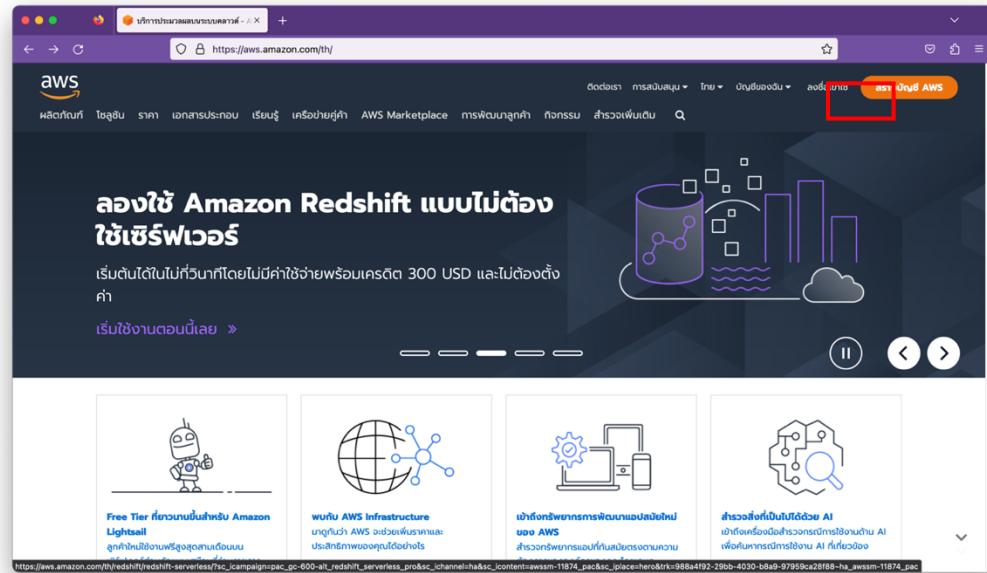
4.6 เมื่อ Import ข้อมูลสำเร็จจะได้ข้อมูลอยู่ใน Collection



ภาพที่ 83 หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass

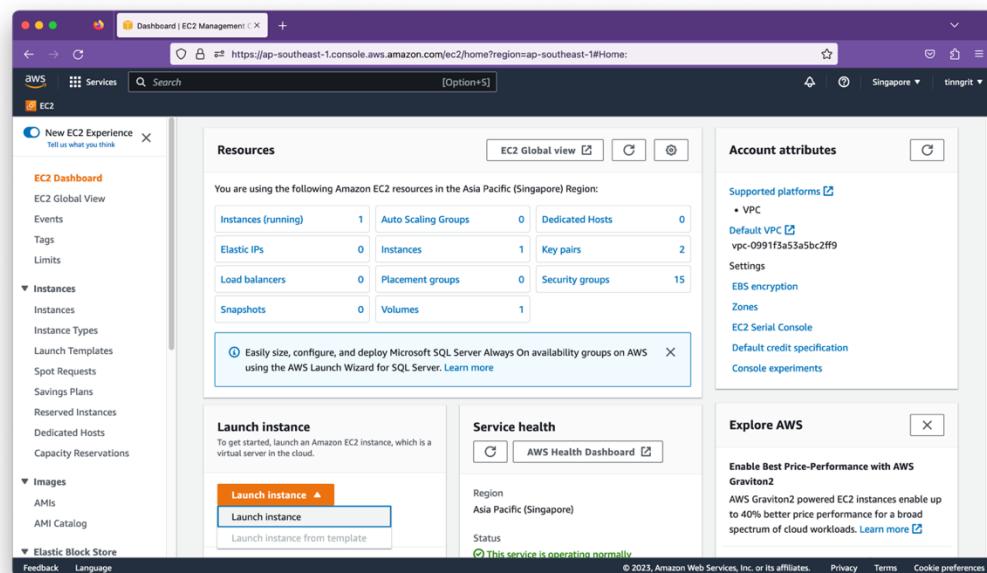
## 5. การใช้งาน Cloud computing ของ Amazon Web Services

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



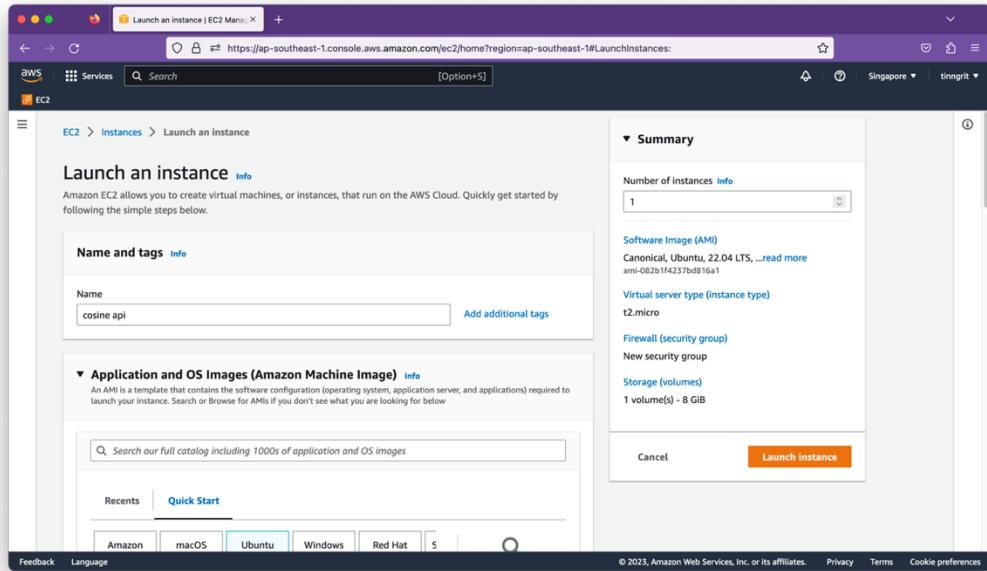
ภาพที่ 84 หน้าเว็บไซต์ Amazon Web Services

5.2 กดที่เมนู EC2 และกดปุ่ม Launch instance เพื่อสร้าง Instance ใหม่



ภาพที่ 85 หน้าแสดงการเลือกสร้าง Instance ใหม่

### 5.3 ตั้งค่าเครื่อง Instance ตามต้องการและกดปุ่ม Launch instance เพื่อทำการสร้าง



ภาพที่ 86 หน้าแสดงการตั้งค่า Instance

### 5.4 ทำการเชื่อมต่อไปยัง Instance ด้วยวิธี SSH โดยใช้ Terminal

```
ubuntu@ip-172-31-35-185: ~
cd desktop/aws
~/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 87 ตัวอย่างการเชื่อมต่อเข้าไปยัง Instance

## 6. การติดตั้งและใช้งาน Web API สำหรับคำนวณค่าความคล้ายคลึง (Cosine similarity)

6.1 ทำการ Clone project จาก Github ที่ลิงค์ [https://github.com/slapexs/final\\_project.git](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 88 การดาวน์โหลดโปรเจคจาก Github ด้วยคำสั่ง git clone มายัง Instance และใช้คำสั่ง cd final\_project เพื่อเข้าไปยังโฟลเดอร์

6.2 สร้างไฟล์ใหม่ในโฟลเดอร์ตั้งชื่อว่า .env และสร้างตัวแปรชื่อว่า MONGODB\_USERNAME MONGODB\_PASSWORD และ CLUSTER เพื่อเก็บค่าเชื่อมต่อฐานข้อมูล

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

ภาพที่ 89 สร้างไฟล์ใหม่ชื่อ .env และสร้างตัวแปรเพื่อเก็บค่าเชื่อมต่อฐานข้อมูล

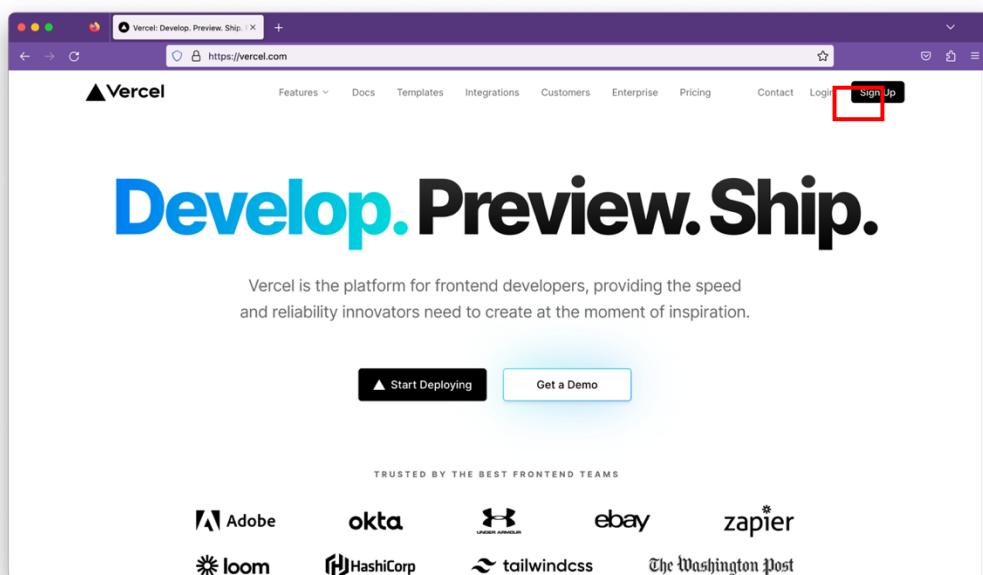
6.3 ใช้คำสั่ง pip3 install –r requirements.txt เพื่อทำการติดตั้ง library ที่จำเป็นและใช้คำสั่ง python3 –m uvicorn api:app เพื่อรัน Server Web API

```
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 90 ตัวอย่างการรัน Server Web API เพื่อคำนวณค่า Cosine similarity บน AWS

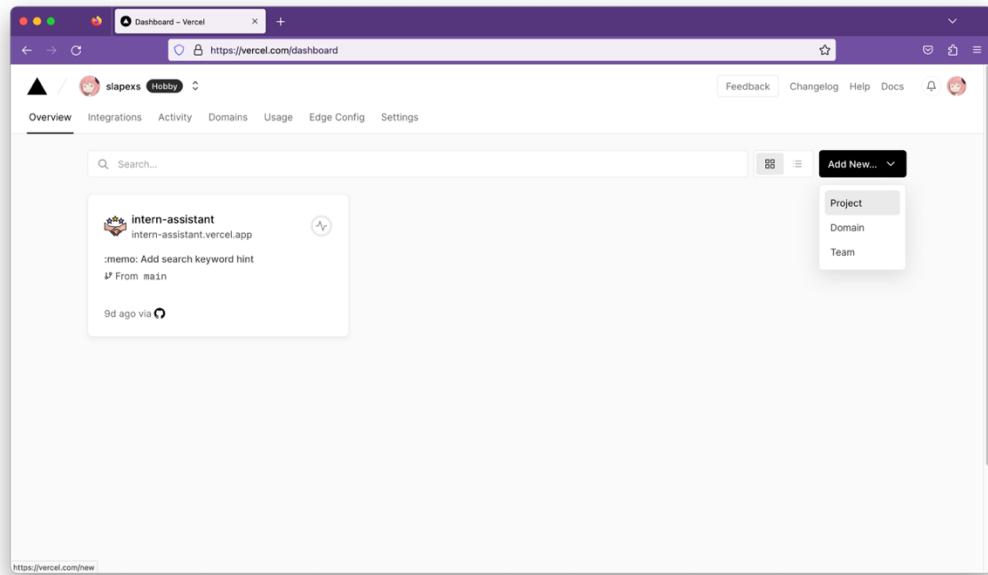
## 7. การติดตั้งและใช้งานเว็บแอปพลิเคชัน (Web application)

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



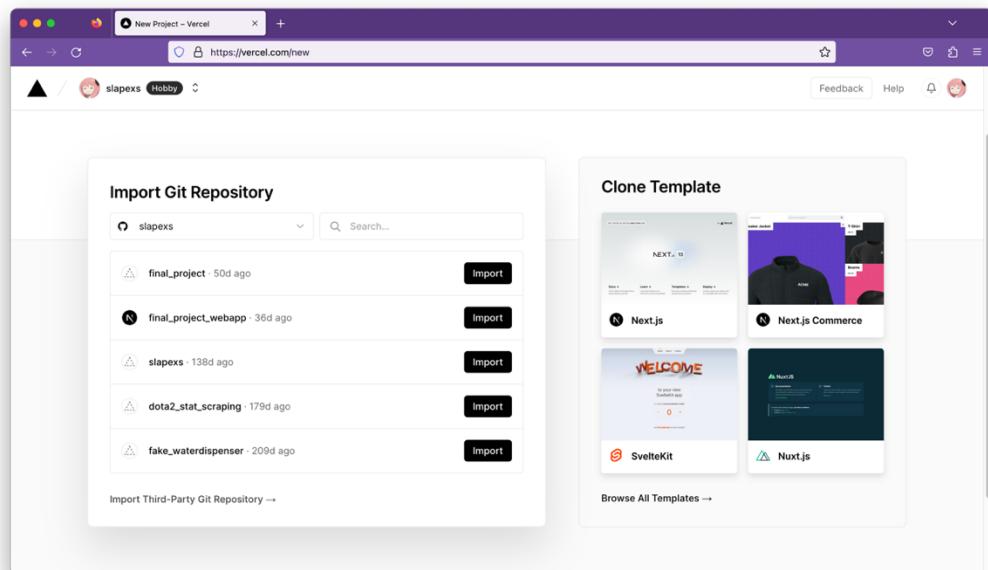
ภาพที่ 91 หน้าเว็บไซต์ Vercel

## 7.2 ทำการสร้างโปรเจคใหม่กดปุ่ม Add New และเลือก Project



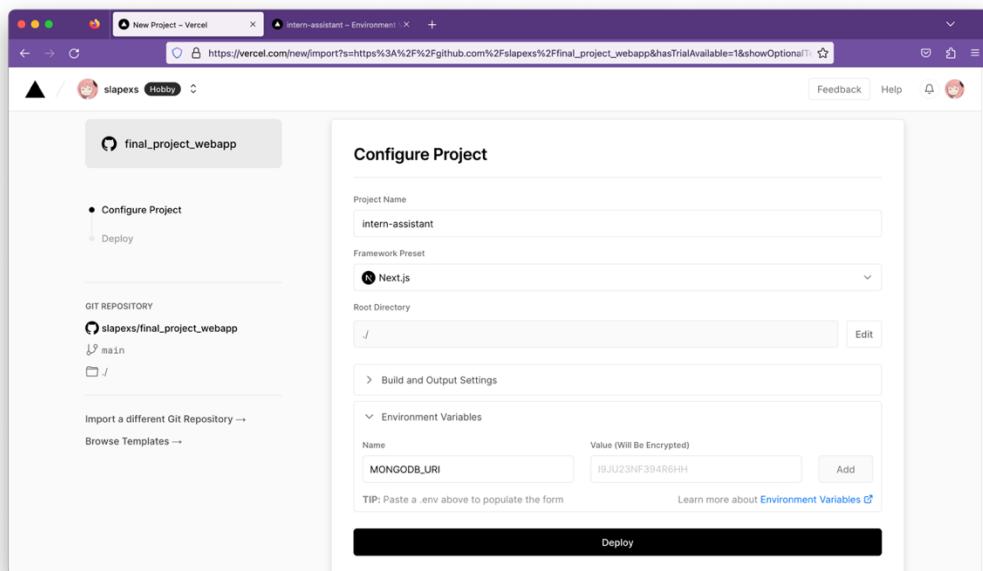
ภาพที่ 92 สร้างโปรเจคใหม่ใน Vercel

## 7.3 เลือก Repository ที่ต้องการจะ deploy และกดปุ่ม Import



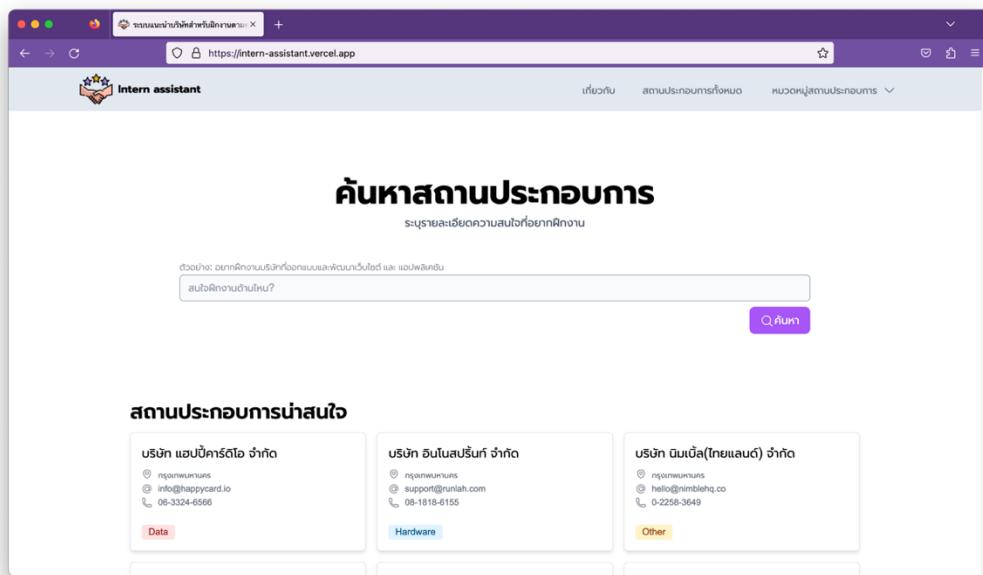
ภาพที่ 93 หน้าแสดงรายชื่อ Repository

7.4 ตั้งค่าโปรเจค และเพิ่มตัวแปร MONGODB\_URI ในส่วนของ Environment Variables และใส่ค่าเป็น Connection string ของ MongoDB Atlas จากนั้นกดปุ่ม Deploy เพื่อทำการเผยแพร่สู่สาธารณะ



ภาพที่ 94 หน้าการตั้งค่าโปรเจคก่อน Deploy

## 7.5 ตัวอย่างหน้าเว็บไซต์เมื่อ Deploy เรียบร้อย



ภาพที่ 95 ตัวอย่างหน้าเว็บไซต์

ภาคผนวก ข  
คู่มือการใช้งาน

## ภาคผนวก ข คู่มือการใช้งาน

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเดเมินท์ (K-Means) สามารถแบ่งได้ 3 ส่วนดังนี้

### 1. ผู้ดูแลระบบ

- 1.1 การจัดกลุ่มข้อมูล
- 1 เปิดไฟล์ clustering.py ในโปรแกรม Text editor
2. ทำการกำหนดจำนวนของกลุ่มข้อมูลที่ตัวแปร k

```

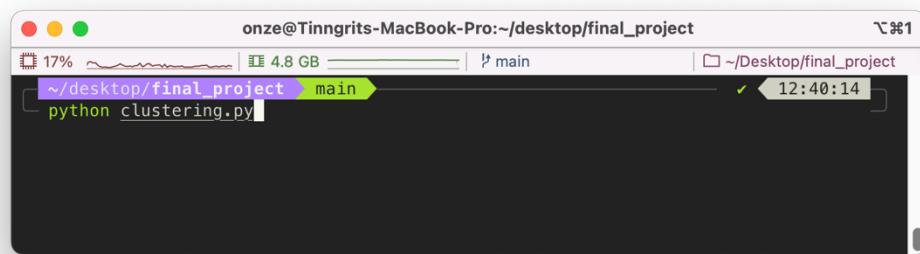
● ● ●

1 k = 7
2 kmeans = KMeans(n_clusters=k, random_state=1)
3 # Fit model
4 kmeans.fit(df_tfidf[['x_value', 'y_value']])
5 clusters = kmeans.labels_

```

ภาพที่ 96 แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k

### 3. ใช้คำสั่งใช้งานไฟล์เพื่อจัดกลุ่มข้อมูล



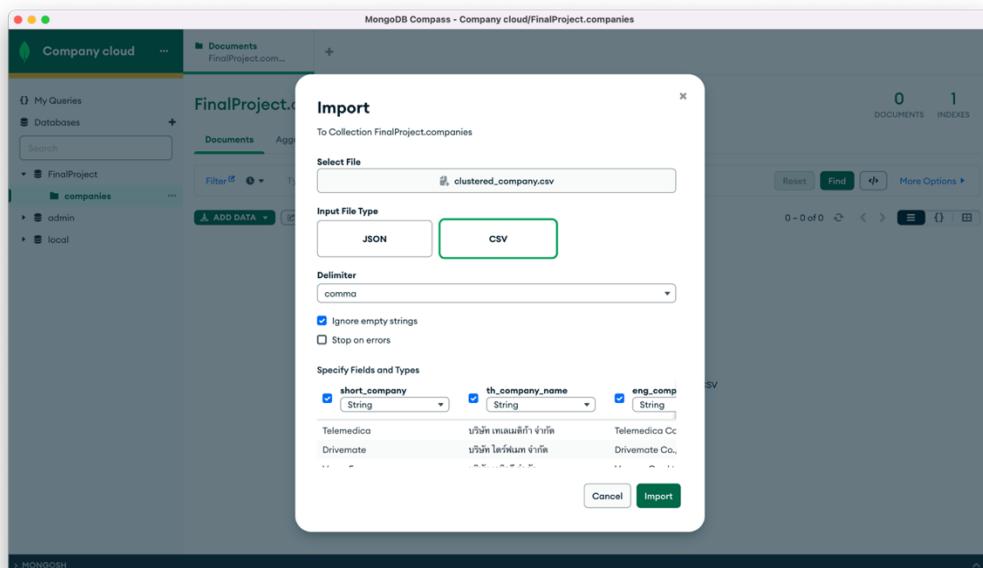
ภาพที่ 97 แสดงการใช้คำสั่งจัดกลุ่มข้อมูลใน Terminal

4. ได้ไฟล์ clustered\_company.csv ในโฟลเดอร์ document ที่เป็นผลลัพธ์การจัดกลุ่มข้อมูล

h_company	Eng_compan	Type_busline	Type_techno	Product	Type_innoval	Detail	Owner	Province_bas	Address	Phone_numb	Email
บริษัท Telematica Co.	SW, Digital Se	เทคโนโลยีสารสนเทศ	มีค่ากับทุกๆ	บริการ Platform	นางสาว ภูริษา	กรุงเทพมหานคร 18,20 ซอยสุขุม 09-0004-000 support@coo...					
บริษัท Driveamate Co.	SW	เทคโนโลยีสารสนเทศ	มีค่ากับทุกๆ	บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 1168 ถนนพหลฯ 0-2026-3238 contact@driv...					
บริษัท เวนิว จำกัด	Venue Co., L	SW, Digital Se	เทคโนโลยีสารสนเทศ	มีค่ากับทุกๆ	บริการ Platform	นาย พิชัยรัชช์	กรุงเทพมหานคร ห้อง 201 อาคาร 0-2120-6013 admin@venu...				
บริษัท ชิปปิ้ง จำกัด	Shippop Co., I	SW, Digital Se	เทคโนโลยีสารสนเทศ	มีค่ากับทุกๆ	บริการ Platform	นางสาว กฤตา พงษ์	กรุงเทพมหานคร 128/403 ชั้น 3, 0-2080-3960 customers@sh...				
บริษัท แอร์พอร์ต	Airportels Co., SW	Digital Se	เทคโนโลยีสารสนเทศ	มีค่ากับทุกๆ	บริการ Platform	นางสาว ภูริษา	กรุงเทพมหานคร 128/39-240 09-2905-3399 contact@ship...				
บริษัท เฟลสตอรี่	Food Stories I	HWNSmart Dev	เทคโนโลยีสารสนเทศ	มีค่ากับทุกๆ	บริการ Platform	นางสาว มนดา	กรุงเทพมหานคร 29/217 หมู่บ้าน 06-3216-6691 center@airpor...				
บริษัท บล็อก จำกัด	Bulk Asia Co., SW				บริการ Platform	นางสาว ภูริษา	กรุงเทพมหานคร 4/11-12 ซอยสุขุม 09-8562-555 info@foodstor...				
บริษัท บล็อก จำกัด	Bulk One Groi	SW			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร ชั้น 14 อาคาร 2 0-2101-2851 support@bul...				
บริษัท เซลสุกิ จำกัด	Selsuki Co., L	SW			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร ชั้น 14 อาคาร 2 0-2101-2851 support@bul...				
บริษัท กอลด์ฟิลด์	Golfdig Co., I	SW, Digital Se			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 5 ซอยสุขุม 4 0 08-9333-1001 hello@golfdi...				
บริษัท อี-แอป จำกัด	I-App Creator	SW, Digital Se			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 23/34-35 อาคาร 08-8668-333 support@app...				
บริษัท เอ็นเนอร์จี้	Energy Respo	HWNSmart Dev, Artificial Intelli			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 408 อาคาร 08-0264-614 contact@ene...				
บริษัท ทัฟ จำกัด	Touchcore Co.	SW, Digital Co	Chatbot		บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 69/5 อาคารที่ 0-240-25626 care@onechat...				
บริษัท ดี.วี.โซลูชัน	Visual Edx Co.	SW			บริการ Platform	นางสาว ใจดี	กรุงเทพมหานคร 264/121 ถนนที่ 08-6419-094 hi@dexi.co...				
บริษัท อินโฟร์ดี	Infored Co.	LT	SW, Digital Co	Artificial Intelli	บริการ Platform	นาย ไชยวัฒน์	กรุงเทพมหานคร 670 ถนนพหลฯ 08-8018-888 contact@inf...				
บริษัท นีโน๊ต!	OneStockton	SW			บริการ Platform	นาย นีโน๊ต!	กรุงเทพมหานคร 1-7 อาคารที่ 0-2026-3223 cs@onestockt...				
บริษัท รีฟิน อินโนเวชัน	Refinn Interna	SW	Artificial Intelli	เทคโนโลยีสารสนเทศ	บริการ Platform	นางสาว มีนาวดา	กรุงเทพมหานคร 897 ซอยสุขุม 0-2080-3924 contact@refin...				
บริษัท ไพริซ่า จำกัด	Prizeca Co., L	SW			บริการ Shopper	นางสาว นิตยาลักษณ์	กรุงเทพมหานคร 126/101 อาคาร 0-2080-6125 info@prizeca...				
บริษัท สกิลเลจ	Skilllane Educ.	SW, Digital Se			บริการ Platform	นางรัตนาพร	กรุงเทพมหานคร 49/5 ถนนเมือง 08-1996-7156 info@skilllan...				
บริษัท โลคาล จำกัด	Local Alike Co	SW			บริการ Platform	นาย สมศักดิ์	กรุงเทพมหานคร 12 ถนนพหลฯ 0-2115-9861 info@localalik...				
บริษัท ไลค์ เม จำกัด	Like Me Co., L	SW			บริการ Platform	นางสาว ใจดี	กรุงเทพมหานคร 216 ถนนสุขุม 09-2264-549 hi@acommore...				
บริษัท ออคบี	Ookbee Co., L	SW, Digital Co	Artificial Intelli	อุตสาหกรรม	บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 1104/207-209 0-2187-2701 member@ook...				
บริษัท จีที.ดี.	JIT Dot Com	SW	Artificial Intelli	เทคโนโลยีสารสนเทศ	บริการ Platform	นางสาว วรรณา	กรุงเทพมหานคร 1111/27 หมู่บ้าน 0-2026-3707 wonder@j...				
บริษัท ฟิกซี่ จำกัด	Fixzy Co., Ltd.	SW, Digital Se			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร 599 ชั้น 3 ถนน 0-2165-0384 info@fixzy.net...				
บริษัท โซเชียล จำกัด	SocialMotion	SW, Digital Co			บริการ Platform	นางสาว อรุณรัตน์	กรุงเทพมหานคร ชั้น 3 ถนน 0-2458-262 hello@sociali...				

ภาพที่ 98 แสดงไฟล์ clustered\_company.csv

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



ภาพที่ 99 หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv

## 6. เมื่อ Import ข้อมูลสำเร็จจะได้ข้อมูลอยู่ใน Collection

The screenshot shows the MongoDB Compass interface with the 'FinalProject.companies' collection selected. The collection contains 1643 documents. Two specific documents are highlighted:

```

{
  "_id": "ObjectId('63fb38ceeb660807118c0972')",
  "short_company": "Teledmedica",
  "th_company_name": "บริษัท เทเลเมดิค้า จำกัด",
  "eng_company_name": "Teledmedica Co., Ltd.",
  "type_business": "SM, Digital Service",
  "product": "แพลตฟอร์มสุขภาพ",
  "type_innovation": "นวัตกรรม",
  "detail": "บริการ Platform ช่วยเหลือให้ผู้คนเข้ามารักษาสุขภาพ ผ่านแอปฯ",
  "owner": "นางสาว อรุณรัตน์ ใจดี",
  "province": "กรุงเทพมหานคร",
  "address": "15, ถนนพญาไท กรุงเทพมหานคร ประเทศไทย",
  "phone_number": "099-0004-9999",
  "email": "support@ooca.co",
  "website": "http://www.ooca.co",
  "source": "http://www.mhesi.go.th/home/index.php/pr/news/1493-4-steps-ooca, http://",
  "cluster": "0"
}

{
  "_id": "ObjectId('63fb38ceeb660807118c0973')",
  "short_company": "Drivemate",
  "th_company_name": "บริษัท ไดรฟ์เมท จำกัด",
  "eng_company_name": "Drivemate Co., Ltd.",
  "type_business": "SM, Peer-to-peer",
  "product": "แพลตฟอร์มสื่อสาร, คอมมูนิตี้",
  "type_innovation": "นวัตกรรม",
  "detail": "บริการ Platform แอร์/เช่า รถ peer-to-peer"
}

```

ภาพที่ 100 หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass

## 2. ผู้ใช้งาน

2.1 เข้าเว็บไซต์ <https://intern-assistant.vercel.app> จากนั้นทำการค้นหาบริษัทด้วยความสนใจในรูปแบบงานของผู้ใช้งาน

The screenshot shows the 'Intern assistant' web application. The main page has a search bar with the placeholder text 'ค้นหาสถานประกอบการ' and a purple 'ค้นหา' button. Below the search bar, there is a section titled 'สถานประกอบการที่สนใจ' (Businesses of interest) with six cards:

- บริษัท ไทย ดอก ค้อน จำกัด
- บริษัท ขายล็อพ จำกัด
- บริษัท มันเน่เก็บ จำกัด
- บริษัท ชอร์กเกอร์ จำกัด
- บริษัท แอบบีซี จำกัด
- บริษัท เอ็กซากอน แอบบ์ จำกัด

ภาพที่ 101 หน้าเว็บไซต์ intern-assistant.vercel.app

## 2.2 เมื่อเจอรายชื่อบริษัทผู้ใช้สามารถกดดูข้อมูลบริษัทได้ตามต้องการ

ภาพที่ 102 หน้าแสดงผลลัพธ์เมื่อค้นหาบริษัท

## 2.3 ผู้ใช้สามารถดูข้อมูลบริษัทเพื่อประกอบการตัดสินใจในการเลือกสถานประกอบการเพื่อฝึกงานได้

ภาพที่ 103 หน้าแสดงข้อมูลบริษัท

### 3. Web API

เป็นเว็บ API ที่ทำหน้าที่ให้ข้อมูลที่ต้องการและมีหน้าที่ในการคำนวณค่าความคล้ายคลึง (Cosine similarity) ข้อมูลที่สามารถคืนไปยังคำขอได้มีดังนี้

### 3.1 ข้อมูลบริษัททั้งหมดในฐานข้อมูล

ภาพที่ 104 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/allcompanies>

จากภาพที่ 99 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.teceh/allcompanies> ด้วยเมธอด GET เพื่อขอข้อมูลบริษัททั้งหมดในฐานข้อมูลด้วยโปรแกรม Postman

### 3.2 ข้อมูลบริษัทที่อยู่ในกลุ่มที่กำหนด

The screenshot shows the Postman application interface. On the left, there's a sidebar with sections for 'Collections', 'APIs', 'Environments', 'Mock Servers', 'Monitors', 'Flows', and 'History'. The main area displays a collection named 'My Workspace' containing several API definitions:

- Calculate cosine similarity**:
  - POST calculate cosine similarity**
  - GET Get company in cluster**
  - POST searchcompany**
  - GET https://lamonze.tech/api**
- web scraping**

A specific API definition for 'Get company in cluster' is selected. The details pane shows the following:

- Method**: GET, URL: <https://lamonze.tech/company/1>
- Params**:
  - Authorization
  - Headers (6)
  - Body
  - Pre-request Script
  - Tests
  - Settings
- Query Params**:

KEY	VALUE	DESCRIPTION
Key	Value	Description
- Body**:
  - Pretty
  - Raw
  - Preview
  - Visualize
  - JSON

```
1 {
2   "short_company": "SOFTLAVU",
3   "th_company_name": "บริษัท ลัมโซน จำกัด",
4   "eng_company_name": "Softlau Co., Ltd.",
5   "type_business": "MW, SW",
6   "detail": "บริษัทฯ เป็นแพลตฟอร์ม (Online Platform - Web/Mobile Application) สำหรับจัดการธุรกิจ (Business Management Software) อาทิเช่น LavuPOS (Lavu's Point of Sale System)",
7   "owner": "มร.วิวัฒน์ ภานุสิน ผู้ก่อตั้งบริษัท",
8   "province_base": "กรุงเทพมหานคร",
9   "address": "ชั้น 1 ถนนสุขุมวิท 59 แขวงคลองเตย เขต 59/19 หมู่ 2 แขวงคลองเตย เขต คลองเตย กรุงเทพมหานคร 10540",
10  "phone_number": "0-2105-4443 ext 1",
11  "email": "sale.softlau@gmail.com",
12  "website": "http://softlau.com/th",
13  "cluster": "1",
14  "company_id": "e11b9783-bb6f-11ed-a5cf-1831bf2b91e6",
15  "cluster_name": "Other"
16 },
17 }
18
19   "short_company": "Wyn",
```

ภาพที่ 105 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/cluster/1>

จากภาพที่ 100 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.teceh/cluster/1> ด้วยเมธอด GET เพื่อขอข้อมูลบริษัทที่อยู่ในกลุ่มที่ 1 ทั้งหมด ในฐานข้อมูลด้วยโปรแกรม Postman

### 3.3 ขอรายการรหัสของกลุ่ม (Cluster ID) ทั้งหมดที่มีในฐานข้อมูล

```

1
2   "clusters": [
3     "0",
4     "1",
5     "2",
6     "3",
7     "4",
8     "5",
9     "6"
10   ]
11

```

ภาพที่ 106 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/cluster>

จากภาพที่ 101 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.teceh/cluster> ด้วยเมธอด GET เพื่อขอข้อมูลรายการรหัสของกลุ่มข้อมูล (Cluster ID) ทั้งหมด ในฐานข้อมูลด้วยโปรแกรม Postman

### 3.4 การส่งค่าเพื่อคำนวณค่าความคล้ายคลึง (Cosine similarity)

#### 1. ส่งกลับข้อมูลมาเป็นผลลัพธ์ค่าความคล้ายคลึงและรหัสกลุ่ม (Cluster ID)

```

1
2   "keywords": "รายงานความเสี่ยงที่เกี่ยวกับเศรษฐกิจโลก และ ผลกระทบทางการเมือง"
3
4
5
6
7
8
9
10
11   "cosine_similarity": [
12     {
13       "0": 0.03215394873323255,
14       "1": 0.0,
15       "2": 0.15109866213365025,
16       "3": 0.18426466647099922,
17       "4": 0.02395706450811518,
18       "5": 0.3040331445123691,
19       "6": 0.03532800669549177
20     ],
21   "max_cosine_similarity": 0.3040331445123691,
22   "cluster": 5
23
24

```

ภาพที่ 107 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/search>

จากภาพที่ 102 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/search> และแนบข้อมูลรายละเอียดความสนใจของผู้ใช้ไปด้วยเพื่อคำนวณค่าความคล้ายคลึง (Cosine similarity) ด้วยเมธอด POST และผลลัพธ์ที่คืนค่ากลับมาจะเป็นค่าความคล้ายคลึงของแต่ละกลุ่ม และรหัสกลุ่มข้อมูลที่มีความคล้ายคลึงมากที่สุด

## 2. ส่งกลับข้อมูลมาเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่มีความคล้ายมากที่สุด

```

{
  "short_company": "MANINNOVATION",
  "th_company_name": "บริษัท มนต์นวัตกรรม จำกัด",
  "eng_company_name": "Manin Innovation Co., Ltd.",
  "type_business": "SM, Digital Content",
  "detail": "เว็บไซต์และแอปพลิเคชัน (Web Application), ออกแบบตราพิเศษ (Logo Design), ออกแบบเอกสาร (Brochure & Stationery Design)",
  "owner": "นายสมชาย วงศ์สมบูรณ์, นายกรุง กิตติพันธุ์, นายปรัชญ์ อรุณพันธุ์",
  "province_base": "กรุงเทพมหานคร",
  "address": "38 ถนนพหลโยธิน 16 แขวงลาดพร้าว เขตลาดพร้าว กรุงเทพมหานคร 10520",
  "phone_number": "0-2727-7274",
  "email": "sm@maninnovation.com",
  "website": "http://thumweb.manininnovation.com",
  "cluster": "5",
  "company_id": "e11b799b-bbf1-11ed-b954-1831bf2b91e6",
  "cluster_name": "Network"
},
{
  "short_company": "Deeboon",
}

```

## ภาพที่ 108 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/searchcompany>

จากภาพที่ 103 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/searchcompany> และแนบข้อมูลรายละเอียดความสนใจของผู้ใช้ไปด้วยเพื่อคำนวณค่าความคล้ายคลึง (Cosine similarity) มีลักษณะคล้ายกับการทำงานในภาพที่ 102 แต่ผลลัพธ์ที่คืนค่ากลับมาจะเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่มีความคล้ายมากที่สุด

## ประวัติผู้ศึกษา



ชื่อ-นามสกุล : นายทินกรฤทธิ์ สิงห์แก้ว

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