



โครงการทางวิทยาการคอมพิวเตอร์ระดับปริญญาตรี

เรื่อง

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่ม
เคลื่อน (K-Means)

COMPUTER INTERNSHIP RECOMMENDATION SYSTEM WITH
K-MEANS CLUSTERING

โดย

นายทินกฤต สิงห์แก้ว

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน
ปีการศึกษา 2565



ใบรับรองโครงการวิทยาการคอมพิวเตอร์
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน^น
วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)

เรื่อง ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means)
Computer Internship Recommendation System With K-Means Clustering
โดย นายทินกฤต สิงห์แก้ว

คณะกรรมการพิจารณาเห็นชอบโดย

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

อาจารย์ที่ปรึกษาร่วม..... วันที่...../...../.....
(อาจารย์วรวิทย์ ผึ้นคำอ้าย)

อาจารย์ผู้รับผิดชอบบวิชา..... วันที่...../...../.....
(อาจารย์ปกรณ์ สุนทรเมธ)

ประธานหลักสูตร..... วันที่...../...../.....
(อาจารย์วรวิทย์ ผึ้นคำอ้าย)

โครงการทางวิทยาการคอมพิวเตอร์ระดับปริญญาตรี

เรื่อง

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่ม
เคลื่อน (K-Means)

Computer Internship Recommendation System With K-Means Clustering

โดย

นายทินกฤต สิงห์แก้ว

คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร
มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน
เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรบัณฑิต สาขาวิทยาศาสตร์
พ.ศ. 2565

บทคัดย่อ

ชื่อโครงการ	: ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัด กลุ่มเคลื่อน (K-Means)
อาจารย์ที่ปรึกษา	Computer Internship Recommendation System With K-Means Clustering
อาจารย์ที่ปรึกษาร่วม	: ผู้ช่วยศาสตราจารย์ ดร. นงนุช เกตุยิ่ง
สาขาวิชา	: อาชญากรรมวิทย์ ผู้นำคำข้าม
หลักสูตร	: วิทยาศาสตร์
ปีการศึกษา	: วิทยาการคอมพิวเตอร์
	: 2565

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) (Computer Internship Recommendation System With K-Means Clustering) เป็นระบบที่ช่วยแนะนำสถานประกอบการสำหรับการฝึกงานของนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน ในรูปแบบของเว็บแอปพลิเคชันที่พัฒนาโดยใช้ Next.js เว็บเฟรมเวิร์ค (Web Framework) สำหรับพัฒนาเว็บแอปพลิเคชัน โดยให้นักศึกษาระบุรายละเอียดความสนใจตามรูปแบบธุรกิจ หรือรูปแบบของงานเพื่อนำมาวิเคราะห์หาความคล้ายคลึงกับข้อมูลสถานประกอบการที่มีอยู่ในฐานข้อมูล ที่ผ่านกระบวนการจัดกลุ่มข้อมูลเคลื่อน (K-Means) ซึ่งเป็นส่วนหนึ่งของเทคโนโลยีปัญญาประดิษฐ์ โดยการใช้เทคโนโลยีประมวลผลภาษาธรรมชาติ (Natural Language Processing) เพื่อจัดกลุ่มข้อมูลสถานประกอบการด้านคอมพิวเตอร์จากสมาคมปัญญาประดิษฐ์แห่งประเทศไทย ซึ่งการจัดกลุ่มข้อมูลด้วยเคลื่อน (K-Means) อยู่ในกลุ่มของการให้คอมพิวเตอร์เรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) โดยผลการทดลองใช้พบว่า นักศึกษาได้ใช้ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) (Computer Internship Recommendation System With K-Means Clustering) และได้มีประสิทธิภาพที่ค่าเฉลี่ยค่าความพึงพอใจอยู่ที่ 4.01 ซึ่งอยู่ในระดับดี

กิจกรรมประจำ

โครงการทางวิทยาการคอมพิวเตอร์ “ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) (Computer Internship Recommendation System With K-Means Clustering)” เพื่อการสำเร็จการศึกษาระดับปริญญาตรี สามารถดำเนินการจนประสบความสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาและคำแนะนำจากคณาจารย์หลาย ๆ ท่านในหลักสูตรวิทยาการคอมพิวเตอร์ ที่ได้กรุณาว่าให้ความรู้ แนะนำแนวทาง ข้อคิด ข้อแนะนำสู่ความสำเร็จ และช่วยแก้ไขปัญหาอุปสรรคต่าง ๆ รวมทั้งรูปเล่มให้สำเร็จไปได้ด้วยดี

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. นงนุช เกตุย และอาจารย์วรวิทย์ พันคำข่าย อาจารย์ที่ปรึกษา และอาจารย์ประจำวิชาทุกท่าน ผู้ชี้งกรุณาว่าให้ความรู้ คำแนะนำแนวทางการสร้างผลงานสู่ความสำเร็จ และช่วยแก้ไขปัญหาอุปสรรคต่าง ๆ รวมทั้งตรวจทานแก้ไขรูปเล่มจนเสร็จสมบูรณ์ ขอขอบคุณบิดา แมรดา ผู้มีพระคุณทุกท่าน เพื่อนนักศึกษา และบุคคลที่เกี่ยวข้องที่ยังไม่ได้กล่าวถึง ที่ได้ช่วยออกแบบความคิดเห็น ได้ให้ข้อแนะนำ และอำนวยความสะดวกในด้านต่าง ๆ ในการทำโครงการครั้งนี้ได้ ณ ที่นี่

สุดท้ายนี้ ผู้ศึกษาหวังว่า โครงการฉบับนี้ จะเป็นประโยชน์สำหรับมหาวิทยาลัย และนักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าจะเป็นประโยชน์ที่จะศึกษาต่อไป

สารบัญ

	หน้า
บทคัดย่อ	๑
กิตติกรรมประกาศ	๒
สารบัญ	๓
สารบัญตาราง	๔
สารบัญภาพ	๕
บทที่ 1 บทนำ	๖
1.1 ความเป็นมาและความสำคัญของปัญหา	๖
1.2 วัตถุประสงค์	๗
1.3 ขอบเขตของโครงการ	๗
1.4 ประโยชน์ที่คาดว่าจะได้รับ	๙
1.5 อุปกรณ์เครื่องมือที่ใช้ในโครงการ	๙
1.6 นิยามศัพท์ที่ใช้ในโครงการ	๑๐
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	๑๑
2.1 ทฤษฎีและความรู้ที่เกี่ยวข้อง	๑๑
2.2 งานวิจัยที่เกี่ยวข้อง	๑๘
บทที่ 3 ขั้นตอนการดำเนินงาน	๒๑
3.1 การเตรียมและวิเคราะห์ข้อมูล	๒๑
3.2 การทำงานของระบบ	๒๘
3.3 การวิเคราะห์และออกแบบระบบ	๒๘
3.4 การออกแบบฐานข้อมูล	๓๕
3.5 การออกแบบหน้าจอ	๓๗
3.6 การใช้งานระบบ	๔๑
บทที่ 4 ผลการดำเนินงาน	๔๒
4.1 การวิเคราะห์และการตัดคำ (Word segmentation)	๔๒
4.2 ขั้นตอนการใช้งานสำหรับผู้ใช้งาน	๕๒
4.3 การวัดค่าความคล้ายคลึง	๕๖
บทที่ 5 สรุปผลการดำเนินงาน	๕๘
5.1 สรุปผลการดำเนินงาน	๕๘
5.2 สรุปปัญหาที่เกิดระหว่างการดำเนินงาน	๕๙
5.3 แนวทางพัฒนาระบบในอนาคต	๕๙

สารบัญ (ต่อ)

	หน้า
5.4 แบบประเมินความพึงพอใจของผู้ใช้	59
เอกสารอ้างอิง	61
ภาคผนวก ก คู่มือการติดตั้งระบบ	64
1. การจัดกลุ่มข้อมูลด้วยเทคนิคทางปัญญาประดิษฐ์เคมีน (K-Means)	65
2. การสร้างคลัสเตอร์ (Cluster) ของฐานข้อมูลมองโกรดีบี (MongoDB) บนเว็บไซต์	67
3. การสร้างฐานข้อมูลในโปรแกรมมองโกรดีบี (MongoDB)	69
4. การนำข้อมูลเข้าสู่ฐานข้อมูลมองโกรดีบี (MongoDB)	71
5. การใช้งาน Cloud computing ของ Amazon Web Services	74
6. การติดตั้งและใช้งาน Web API สำหรับคำนวณความคล้ายคลึง (Cosine similarity)	76
7. การติดตั้งและใช้งานเว็บแอปพลิเคชัน (Web application)	78
ภาคผนวก ข คู่มือการใช้งาน	80
1. ผู้ดูแลระบบ	81
2. ผู้ใช้งาน	83
3. Web API	85
ประวัติผู้ศึกษา	87

สารบัญตาราง

ตารางที่	หน้า
1. ตัวอย่างการคำนวณค่า Term Frequency ที่จำนวนคำทั้งหมดเท่ากับ 7	7
2. ตัวอย่างการคำนวณค่า Inverse Document Frequency ที่จำนวนเอกสารเท่ากับ 10	8
3. ตัวอย่างการคำนวณค่า TF-IDF	8
4. การวิเคราะห์ข้อมูล	21
5. คำอธิบาย Use case คู่มือการใช้งาน	29
6. คำอธิบาย Use case ดูรายชื่อบริษัททั้งหมด	29
7. คำอธิบาย Use case ดูรายชื่อบริษัทในกลุ่มทั้งหมด	30
8. คำอธิบาย Use case ดูข้อมูลบริษัท	30
9. คำอธิบาย Use case คนหาบริษัทด้วยความสนใจ	31
10. คำอธิบาย Use case แก้ไขคู่มือการใช้งาน	31
11. คำอธิบาย Use case เพิ่ม ลบ และแก้ไขข้อมูลบริษัท	31
12. อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การค้นหาบริษัทด้วยความสนใจ	33
13. อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มใหม่	34
14. พจนานุกรมข้อมูลบริษัท	36
15. ตารางตัวอย่างการวัดค่าความแม่นยำในการตัดคำ	44
16. ผลการทดสอบความแม่นยำของการตัดคำ	45
17. แสดงการนับจำนวนบริษัทแต่ละประเภทในการจัดกลุ่มทั้งหมด 1,643 รายการ	49
18. แสดงจำนวนค่าเฉลี่ยความพึงพอใจต่อระบบ	60

สารบัญภาพ

ภาพที่	หน้า
1. ตัวอย่างการทำ Word segmentation	7
2. การกำหนดสุ่มกำหนดจุด Centroid	9
3. จุด Centroid ที่อยู่ตรงกลางและจุดข้อมูลทุกจุดไม่เปลี่ยนแปลง	10
4. ตัวอย่างการลุ่มข้อมูลที่มีจุด Centroid เป็นภาคบาทสี่เหลี่ยม	10
5. กราฟที่แสดงจำนวนของผิดพลาดเพื่อหาจำนวนกลุ่มที่เหมาะสมที่สุด	11
6. ตัวอย่างโค้ดสำหรับการสร้าง Web API ด้วย fastAPI	13
7. ผลลัพธ์แสดงคำว่า Hello project จาก fastAPI	13
8. ตัวอย่างข้อมูลแบบ JSON	15
9. การเตรียมและวิเคราะห์ข้อมูล	22
10. การเรียกใช้ไลบรารี (Library) สำหรับคำนวณค่า TF-IDF	23
11. การอ่านข้อมูลจากไฟล์และกำหนดตัวกรองการตัดคำ	23
12. พังก์ชันสำหรับใช้ลบตัวเลข และอักขระพิเศษ	24
13. พังก์ชันสำหรับใช้ลบคำที่ไม่สื่อความหมายและตัวเลขไทย	24
14. การวนซ้ำข้อมูลเพื่อตัดคำและทำความสะอาดข้อมูล	24
15. การเหรอและทำการทดสอบโมเดลการคำนวณค่า TF-IDF	25
16. ผลลัพธ์การคำนวณค่า TF-IDF	25
17. การเรียกใช้ไลบรารีสำหรับการจัดกลุ่มข้อมูลด้วยเคมีน (K-Means)	26
18. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล	26
19. การเรียกใช้งานไฟล์ clustering.py เพื่อจัดกลุ่มข้อมูลและบันทึกผลลัพธ์	27
20. การนำเข้าข้อมูลลงสู่ฐานข้อมูล MongoDB	27
21. การทำงานของระบบ	28
22. Use Case Diagram ของระบบ	29
23. Sequence Diagram การคุนหาบริษัทด้วยความสนใจของผู้ใช้	32
24. Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบริษัทใหม่	33
25. Activity Diagram ของผู้ใช้งาน	34
26. Activity Diagram ของผู้ดูแลระบบ	35
27. ER Diagram ระบบแนะนำบริษัทสำหรับฝึกงานตามความสนใจ	36
28. หน้าแรก	38
29. หน้าเกี่ยวกับ	38
30. หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด	39
31. หน้าแสดงรายชื่อบริษัททั้งหมด	39
32. หน้าแสดงผลลัพธ์รายชื่อบริษัท	40
33. หน้าแสดงข้อมูลบริษัท	40
34. ตัวอย่างผลลัพธ์จากการค้นหาด้วยความสนใจของผู้ใช้	41
35. ตัวอย่างข้อมูลต้นฉบับ	43

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
36. ตัวอย่างการตัดคำโดยใช้ Engine newmm	43
37. ตัวอย่างการตัดคำโดยใช้ Engine longest	43
38. ตัวอย่างการตัดคำโดยใช้ Engine deepcut	44
39. ผลการวัดค่าความแม่นยำในการตัดคำของ Engine ในไลบรารี Pythainlp	45
40. ตัวอย่างตาราง TF-IDF แสดงน้ำหนักของคำ	46
41. ตัวอย่างการตัดคำและลบ Stop word	46
42. การทำ Elbow method	47
43. จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม	47
44. จัดกลุ่มข้อมูลจำนวน 8 กลุ่ม	48
45. จัดกลุ่มข้อมูลจำนวน 7 กลุ่ม	48
46. จัดกลุ่มข้อมูลจำนวน 6 กลุ่ม	48
47. เปรียบเทียบอัตราการเติบโตการจัดกลุ่มข้อมูล	50
48. หน้าแรกเว็บไซต์ Intern-assistant	52
49. คนหาบริษัท	53
50. หน้าแสดงผลลัพธ์การค้นหา	53
51. หน้าเกี่ยวกับ	54
52. หน้าแสดงรายชื่อบริษัททั้งหมด	54
53. หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด	55
54. หน้ารายละเอียดบริษัท	55
55. ตัวอย่างการคำนวณค่า Cosine similarity	56
56. ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าความคล้ายคลึง	56
57. ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่คล้ายที่สุด	57
58. โค้ดคำสั่งในไฟล์ clustering.py ใช้ในการจัดกลุ่มข้อมูล	65
59. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล	66
60. แสดงการใช้งานคำสั่งจัดกลุ่มข้อมูลใน Terminal	66
61. แสดงไฟล์ clustered_company.csv	67
62. แสดงหน้าการจัดการ Cluster MongoDB	67
63. แสดงหน้าตั้งค่าและสร้าง Cluster MongoDB	68
64. แสดงหน้าสร้างบัญชีสำหรับจัดการฐานข้อมูล	68
65. แสดงหน้าเพิ่ม IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้	69
66. แสดงหน้าจัดการ Cluster MongoDB	69
67. แสดงการสร้างฐานข้อมูล MongoDB	70
68. แสดงหน้าต่างการสร้างฐานข้อมูลและ Collection	70
69. ตัวอย่างการเลือกตั้งค่าการดาวน์โหลดโปรแกรม MongoDB compass	71
70. หน้าต่างเลือกเชื่อมต่อกับ Cluster	71

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
71. หน้าต่างข้อมูลการเชื่อมต่อ Cluster กับ MongoDB compass	72
72. หน้าต่างโปรแกรม MongoDB compass สำหรับเชื่อมต่อ Cluster	72
73. หน้าต่างโปรแกรมแสดงข้อมูลใน Collection	73
74. หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv	73
75. หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass	74
76. หน้าเว็บไซต์ Amazon Web Services	74
77. หน้าแสดงการเลือกสร้าง Instance ใหม่	75
78. หน้าแสดงการตั้งค่า Instance	75
79. ตัวอย่างการเชื่อมต่อเข้าไปยัง Instance	76
80. การดาวน์โหลดโปรเจคจาก Github ด้วยคำสั่ง git clone	76
81. สร้างไฟล์ใหม่ชื่อ .env และสร้างตัวแปรเพื่อกีบค่าเชื่อมต่อฐานข้อมูล	77
82. ตัวอย่างการรีมัตต์น์ Server Web API เพื่อคำนวนค่า Cosine similarity บน AWS	77
83. สร้างโปรเจคใหม่ใน Vercel	78
84. หน้าแสดงรายชื่อ Repository	78
85. หน้าการตั้งค่าโปรเจคตอน Deploy	79
86. ตัวอย่างหน้าเว็บไซต์	79
87. แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k	81
88. แสดงการใช้คำสั่งจัดกลุ่มข้อมูลใน Terminal	81
89. แสดงไฟล์ clustered_company.csv	82
90. หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv	82
91. ข้อมูลใน Collection ในโปรแกรม MongoDB compass	83
92. หน้าเว็บไซต์ intern-assistant.vercel.app	83
93. หน้าแสดงผลลัพธ์เมื่อค้นหาบริษัท	84
94. หน้าแสดงข้อมูลบริษัท	84
95. ตัวอย่างการส่งคำขอไปยัง https://iamonze.tech/allcompanies	85
96. ตัวอย่างการส่งคำขอไปยัง https://iamonze.tech/company/1	85
97. ตัวอย่างการส่งคำขอไปยัง https://iamonze.tech/cluster	86
98. ตัวอย่างการส่งคำขอไปยัง https://iamonze.tech/search	86
99. ตัวอย่างการส่งคำขอไปยัง https://iamonze.tech/searchcompany	87

บทที่ 1 บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในระบบการศึกษาระดับปริญญาตรีนี้รายวิชาที่มีในการศึกษาปีสุดท้ายของหลักสูตรคือ รายวิชาที่จะต้องให้นักศึกษาแต่ละคนนั้นออกแบบไปฝึกทำงานที่สถานประกอบการต่าง ๆ ในช่วงระยะเวลาหนึ่ง ซึ่งเป็นสิ่งที่มีความสำคัญและความท้าทายเนื่องจากเป็นการที่นักศึกษาจะได้ทดลองทำงานจริง สถานการณ์จริง สถานที่จริง ในสถานประกอบการที่นักศึกษาได้เลือก

ดังนั้นการเลือกสถานประกอบการสำหรับฝึกงานจึงเป็นเรื่องที่ต้องให้ความสำคัญเป็นอย่างมากเนื่องจากหากสถานประกอบการที่เลือกนั้นรูปแบบธุรกิจหรืองานที่ทำนั้น ตรงกันกับความสามารถของนักศึกษา ก็จะเป็นผลดี เนื่องจากความรู้และทักษะที่ได้จากการทำงานนั้นสามารถนำไปต่อยอดและใช้งานจริงเมื่อจบการศึกษาและเข้าทำงาน แต่หากสถานประกอบการที่เลือกนั้น รูปแบบธุรกิจหรืองานที่ทำไม่ตรงกับความต้องการหรือทักษะของนักศึกษาอาจทำให้การฝึกงานนั้นล้มเหลว หรืออาจไม่ได้ความรู้และทักษะที่ต้องการได้ และด้วยเทคโนโลยีปัญญาประดิษฐ์ในด้านของการประมวลผลภาษาธรรมชาตินั้นพัฒนาไปมากขึ้นมาก ทั้งในแง่ของเทคนิค เครื่องมือ และองค์ความรู้ ทำให้เกิดตัวอย่างการนำข้อมูลมาประมวลผลที่มีประสิทธิภาพมากmay ในปัจจุบัน

ด้วยเหตุนี้จึงได้มีการเริ่มโครงการพัฒนาเว็บแอปพลิเคชันที่นักศึกษามีความสนใจในรูปแบบธุรกิจของสถานประกอบการนั้น เพื่อช่วยอำนวยความสะดวกแก่นักศึกษา ให้สามารถนำข้อมูลทางธุรกิจมาใช้ในการตัดสินใจ ในการค้นหาสถานประกอบการสำหรับออกฝึกงาน โดยการใช้วิธีประมวลผลภาษาธรรมชาติเข้ามาช่วยจัดกลุ่มสถานประกอบการและเสนอรายชื่อสถานประกอบการที่เหมาะสมแก่นักศึกษาผ่านทางเว็บแอปพลิเคชัน

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 เว็บแอปพลิเคชันสามารถให้ข้อมูลบริษัทเพื่อการตัดสินใจในการเลือกบริษัทฝึกงานได้

1.3.3 ใช้เทคโนโลยีการประมวลผลภาษาธรรมชาติ เพื่อจัดการคำและแบ่งกลุ่มข้อมูลโดยใช้เทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) และหาความคล้ายของข้อมูลด้วยการคำนวณความคล้าย cosine (Cosine similarity)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 ได้พัฒนาเว็บแอปพลิเคชันสำหรับการค้นหาสถานประกอบการสำหรับผู้คน
- 1.4.2 สามารถนำระบบประมวลผลภาษาธรรมชาติมาใช้ในการจัดกลุ่มข้อมูลได้อย่างแม่นยำ
- 1.4.3 เป็นช่องทางสำหรับการเลือกและหาข้อมูลของสถานประกอบการสำหรับออกผีกงานของนักศึกษา

1.5 อุปกรณ์เครื่องมือที่ใช้ในโครงงาน

1.5.1 Programming 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 นิยามศัพท์ที่ใช้ในโครงการ

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 ประมวลผลโดยที่ไม่ได้จำกัดขอบไว้แต่ให้คอมพิวเตอร์ประมวลผลและกำหนดเองว่าคำตอบควรจะเป็นลักษณะใดบ้าง

1.6.3 การตัดคำ (Word segmentation) หมายถึง ด้วยที่การเขียนภาษาไทยนั้นไม่มีการแยกคำด้วยการเว้นวรรคหรือเส้นภาษาอังกฤษ 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 หรือผู้ใช้อื่น ๆ ใช้ติดต่อสื่อสารเชื่อมต่อแลกเปลี่ยนข้อมูลกัน

1.6.6 การเข้าใช้ระบบคอมพิวเตอร์ (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 การนำความต้องการของผู้ใช้มาเปรียบเทียบกับข้อมูล และคำนวณคาดคะเนความคล้ายคลึง

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาคนคว้าเพื่อจัดทำโครงการทางวิทยาการคอมพิวเตอร์ ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) ผู้ศึกษาได้ศึกษาคนคว้าเอกสารและงานวิจัยที่เกี่ยวข้องตามหัวข้อลำดับต่อไปนี้

2.1 ทฤษฎีและความรู้ที่เกี่ยวข้อง

- 2.1.1 ทฤษฎี การประมวลผลภาษาธรรมชาติ หรือ Natural language processing
- 2.1.2 ทฤษฎี การตัดคำในภาษาไทยหรือ Word segmentation
- 2.1.3 ทฤษฎี การลอกติดความของข้อความด้วยเทคนิค 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 เป็นเทคนิคหนึ่งในเทคโนโลยีปัญญาประดิษฐ์ ที่จะทำให้คอมพิวเตอร์สามารถ

เข้าใจและเรียนรู้ ประมวลผลภาษาของมนุษย์ได้ ในด้านของการวิเคราะห์ภาษาศาสตร์ การตีความจากบทความ หรือการทั้งการแปลภาษา NLP นั้นจำเป็นต้องใช้ความรู้จากหลาย ๆ ศาสตร์เข้ามา เช่น Mathematics, Linguistics, Psychology เพื่อพัฒนาประสิทธิภาพการทำงานและความฉลาดของคอมพิวเตอร์

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

กระบวนการทำงานของ NLP นั้น มีประกอบไปด้วยหลายส่วนของการประมวลผลและใช้แปลความหมาย ประกอบด้วยดังนี้

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

ปัจจุบัน NLP นั้นอยู่ในหลายรูปแบบครอบคลุมกว้างมาก เช่น Digital marketing, ทางการแพทย์ การแปลงภาษา Chatbot และอื่น ๆ (ตาเยะ, 2022)

2.1.2 ทฤษฎี การตัดคำในภาษาไทย (Word segmentation)

การที่นำประโยชน์มาตัดคำออกเป็นคำ ๆ (Word segmentation) เนื่องจากในบางภาษา เช่น ภาษาไทยรูปแบบการเขียนนั้นไม่มีการเว้นวรรคของคำ ต่างจากภาษาอังกฤษที่ใช้การเว้นวรรคในแต่ละคำดังนั้นถ้าจะทำ NLP ที่เป็นภาษาไทยนั้นจำเป็นต้องทำ Word segmentation เพื่อให้ได้ชุดคำที่จะนำไปใช้งานต่อ ในปัจจุบันการทำ Word segmentation นั้นมีเครื่องมือให้ใช้อยู่จำนวนมากยกตัวอย่าง เช่น Python library pythainlp, nltk หรือสามารถใช้บริการ web API ของ aiforthai

```
onzze@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
['น่อน', 'ตาม', 'ดู', 'ดาว']

~/d/final_project main !1 ?3  4s backup_finalproject 14:06:09
```

ภาพที่ 1 ตัวอย่างการทำ Word segmentation

จากภาพที่ 1 เป็นการทำ Word segmentation ด้วย Python library pythainlp จากคำว่า “น่อนตามดูดาว” ได้ผลลัพธ์ออกมาระบุเป็น น่อน, ตาม, ดู, ดาว (L, 2019)

2.1.3 ทฤษฎี การสกัดใจความของข้อความด้วยเทคนิค TF-IDF

การสกัดใจความของข้อความ (Term Frequency – Inverse Document Frequency: TF-IDF) เป็นเทคนิคที่พิจารณาองค์ประกอบของคำภาษาไทยโดยใช้ Term Frequency (TF) และ Inverse Document Frequency (IDF) องค์ประกอบแรก Term Frequency (TF) นั้นหมายถึงการที่หาคำที่มีการใช้ซ้ำบ่อยที่สุดในเอกสารนั้น ๆ ซึ่งแสดงไปถึงว่าคำนั้นเป็นคำที่มีความสำคัญมากของเอกสารนั้น วิธีคำนวณค่าความถี่ของคำใช้การนำจำนวนครั้งของคำที่ปรากฏในเอกสารมาหารด้วยจำนวนคำทั้งหมดในเอกสาร เช่น ต้องการหาค่าความถี่ของคำว่า “เง็บไซต์” ในเอกสาร (CHAKRIT, 2019)

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

ตาราง 1 ตัวอย่างการคำนวณค่า Term Frequency ที่จำนวนคำทั้งหมดเท่ากับ 7

คำ	จำนวนคำ	Term Frequency	ผลลัพธ์
เง็บไซต์	5	$5 \div 7$	0.71
หนังสือ	1	$1 \div 7$	0.14
ออนไลน์	2	$2 \div 7$	0.29
ขาย	2	$2 \div 7$	0.29
เข้าชม	1	$1 \div 7$	0.14
เมื่อถึง	4	$4 \div 7$	0.57
และ	3	$3 \div 7$	0.43

จากตัวอย่างจะเห็นได้ว่าคำว่า “เว็บไซต์” ปรากฏบ่อยในเอกสารทำให้มีค่า Term Frequency สูงจึงเรียกได้ว่าเป็นคำสำคัญของเอกสาร และการใช้ค่า Term Frequency เพื่อหาใจความสำคัญเพียงอย่างเดียวนั้นยังไม่ดีพอ จึงต้องใช้องค์ประกอบ Inverse Document Frequency (IDF) เข้ามาช่วยเพิ่มเติม Inverse Document Frequency (IDF) หมายถึง การคำนวณหน้าหนักของคำโดยการนำคำสำคัญคุ้นหูจากหลาย ๆ เอกสารหากคำนั้นมีค่า Inverse Document Frequency (IDF) ต่ำ แสดงว่าคำนั้นไม่ได้เป็นคำสำคัญของเอกสารทั้งหมด สมการที่ใช้คำนวณหาค่า Inverse Document Frequency (IDF)

$$IDF(\text{ของคำคำหนึ่ง}) = \log\left(\frac{\text{จำนวนเอกสารทั้งหมด}}{\text{จำนวนเอกสารที่มีคำคำนั้นปรากฏ}}\right)$$

ตาราง 2 ตัวอย่างการคำนวณค่า Inverse Document Frequency ที่จำนวนเอกสารเท่ากับ 10

คำ	จำนวนเอกสารที่ปรากฏ	Inverse Document Frequency	ผลลัพธ์
เว็บไซต์	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
เว็บไซต์	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

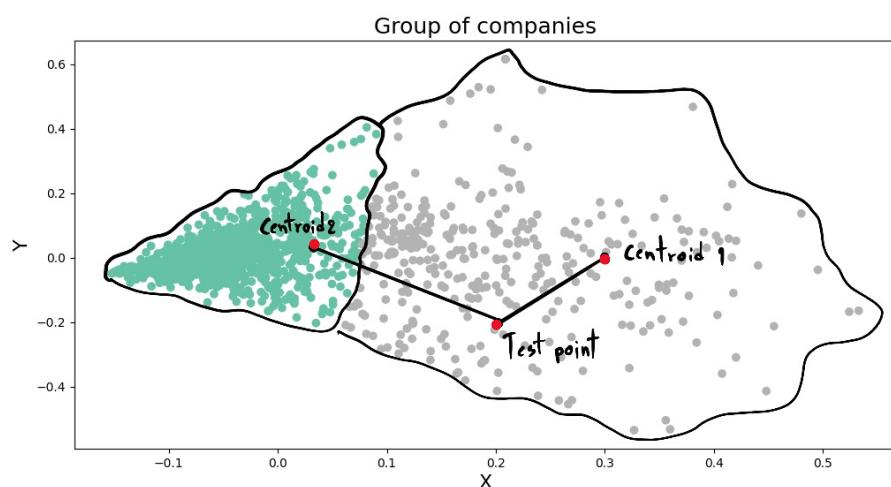
ดังตัวอย่างจะเห็นได้ว่าบางคำที่มีค่า TF-IDF สูงแต่ไม่ได้บ่งบอกถึงลักษณะของข้อความในเอกสาร เช่นคำว่า “และ” ซึ่งถือว่าเป็น Stop word ซึ่งเป็นคำที่ไม่สื่อความหมาย โดยปกติแล้วคำเหล่านี้มักถูกกรองออกก่อนที่จะมีการนำข้อความมาทำการประมวลผลทางภาษา จากตัวอย่างจะเห็นว่าเมื่อคำนวนหาค่า Term Frequency – Inverse Document Frequency (TF-IDF) โดยที่ตัดคำที่ไม่มีความหมายหรือ Stop word ออกแล้วจะเหลือคำว่า “มือถือ” “เว็บไซต์” “ออนไลน์” เรียงลำดับความสำคัญจากมากไปน้อยตามลำดับ (Patipan, 2020)

2.1.4 ทฤษฎี การจัดกลุ่มข้อความด้วยอัลกอริทึม K-Means

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

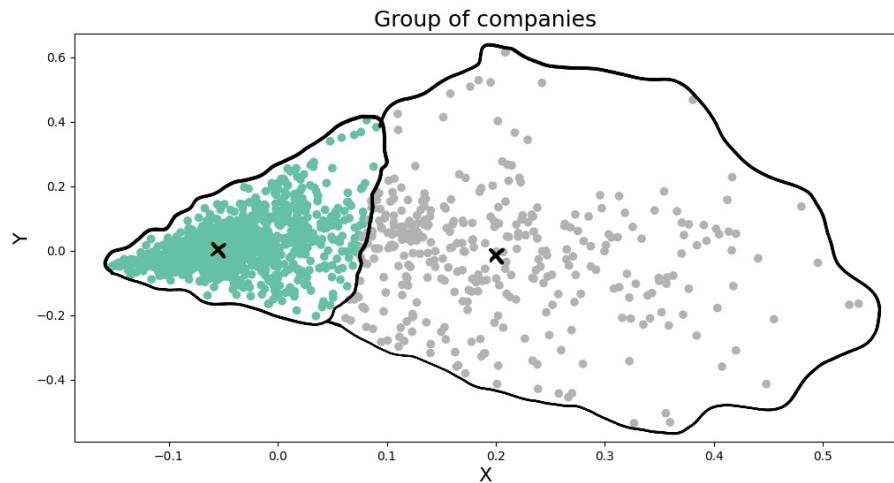
$$Distance = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

ตัวอย่างการคำนวนค่าการกำหนดจุด Centroid เมื่อกำหนด k เท่ากับ 2 ทำการสุมจุดข้อมูลจากข้อมูลทั้งหมด



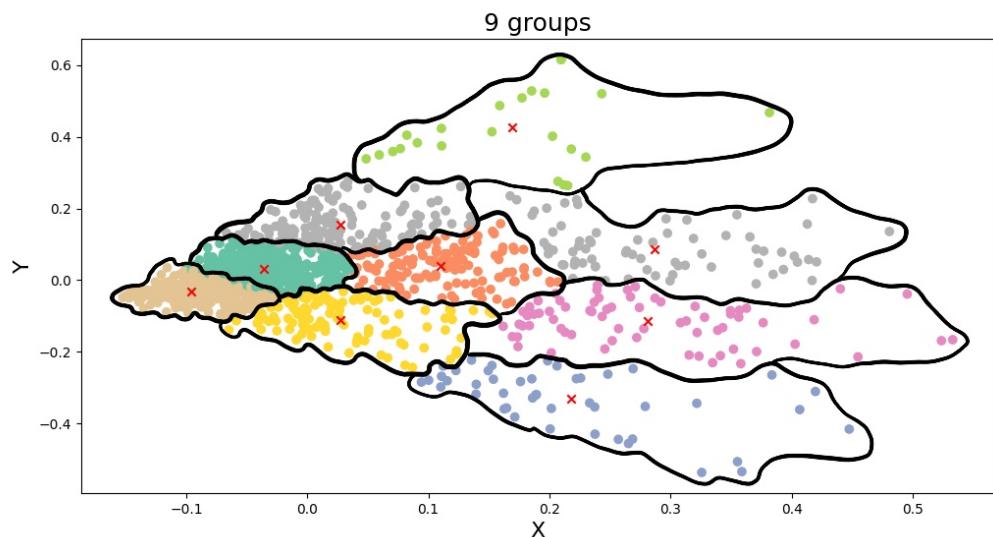
ภาพที่ 2 การกำหนดสุมกำหนดจุด Centroid

ทำการทำซ้ำการกำหนดจุด centroid จนกว่าตำแหน่งของข้อมูลทุกตัวจะไม่เปลี่ยนแปลง



ภาพที่ 3 จุด Centroid ที่อยู่ตรงกลางและจุดข้อมูลทุกจุดไม่เปลี่ยนแปลง

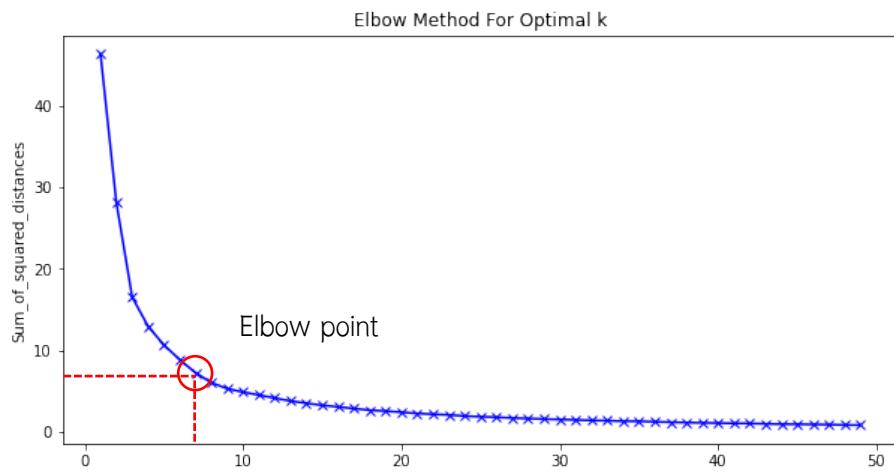
ทำการคำนวณและขยายจุด Centroid และหาค่าเฉลี่ยจนค่าเฉลี่ยไม่มีการเปลี่ยนแปลงจะได้จุดกึ่งกลางของข้อมูลในแต่ละกลุ่ม



ภาพที่ 4 ตัวอย่างการกลุ่มข้อมูลที่มีจุด Centroid เป็นกากบาทสีแดง

2.1.5 ทฤษฎี การหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow method

Elbow method เป็นวิธีหนึ่งที่ใช้หาจำนวนของกลุ่มที่เหมาะสมด้วยการวัดขอผิดพลาด (Error measurement) ผลรวมระยะห่างระหว่างข้อมูลกับจุด Centroid เมื่อขอผิดพลาดน้อยลงความชันของเส้นโค้งจะแบนราบไปตามแกน X จนทำให้เกิดมุมลักษณะเหมือนกับข้อต่อของโครงสร้างของข้อมูลที่เหมาะสมดังในภาพตัวอย่างภาพที่ 5 จะเห็นได้ว่าจำนวนกลุ่มที่เหมาะสมสมควรเป็น 6-7 กลุ่ม (Paul, 2021)



ภาพที่ 5 กราฟที่แสดงจำนวนข้อผิดพลาดเพื่อหาจำนวนกลุ่มที่เหมาะสมที่สุด

2.1.6 ทฤษฎี การคำนวณค่าความคล้ายคลึงด้วยเทคนิค Cosine similarity

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

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

อธิบายโดยง่ายคือเป็นการวัดระยะห่างระหว่าง Object A และ Object B ว่ามีความคล้ายกันแค่ไหน ยกตัวอย่างเบริยบเทียบระหว่างคำว่า “ยินดีที่ได้รู้จักรับ” และ “ยินดีที่ได้รู้จักคะ” ทำการตัดเพื่อหาคำทั้งหมดก่อนคือ “ยินดี” , “ที่” , “ได้” , “รู้จัก” , “รับ” , “คะ” จากนั้นจะได้ Object A และ B เป็นชุดข้อมูล ได้ดังนี้ (Supalerk, 2020)

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

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

$$\text{similarity} = \frac{4}{\sqrt{6} \times \sqrt{6}} = \frac{4}{6} = 0.67$$

$$\text{similarity} = 0.8$$

2.1.7 ทฤษฎี การจัดการระบบคลาวด์ (Amazon web service)

AWS เป็นตัวย่อของ Amazon Web Services ซึ่งเป็นบริการบนระบบคลาวด์ ที่มีบริการหลากหลายมากกว่า 200 โซลูชัน ถูกใช้งานในธุรกิจและองค์กรทุกประเภทไม่ว่าจะเป็นบริษัท สตาร์ทอัป 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 เว็บไซต์ซื้อขายสินค้าออนไลน์ชื่อดังจากประเทศสหรัฐอเมริกา (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 คือกลไกที่ช่วยให้ส่วนประกอบซอฟต์แวร์สองส่วนสามารถสื่อสารกันได้โดยใช้ชุดคำจำกัดความและโปรโตคอล ตัวอย่างเช่น ระบบซอฟต์แวร์ของสำนักพยากรณ์อากาศประกอบด้วยข้อมูลสภาพอากาศรายวัน (API คืออะไร)

2.1.9 ทฤษฎี Cloudflare

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

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

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

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

2.1.10 ទម្រង់ Cors

การอนุญาตการแบ่งปันข้อมูลกัน (Cross-Origin Resource Sharing:CORS) เป็นกลไกที่ใช้เพิ่มเติมเพื่อให้เบราว์เซอร์ได้รับสิทธิ์ในการเข้าถึงทรัพยากรที่เลือกจากเซิร์ฟเวอร์บนโดเมนอื่นมาแสดงบนหน้าเว็บเบราว์เซอร์ได้ คอมพิวเตอร์แต่ละเครื่องต้องมี Protocol ที่เหมือนกัน ถึงจะสื่อสารกันรู้เรื่อง เว็บเบราว์เซอร์จะส่ง HTTP request เมื่อต้องการขอข้อมูลข้ามโดเมนหรือ port ที่ต่างกัน และต้องทำงานของกลางการสื่อสาร (Protocol) เพราบจุบันเรามักจะแยกผัง Front-end และ Back-end ออกจากกันเป็นคนละโดเมน ด้วยเหตุผลเรื่องความปลอดภัยของ Browsers HTTP การอนุญาตให้เข้าถึงแหล่งข้อมูลจะต้องอยู่โดเมนเดียวกันเท่านั้น เว้นแต่ว่าแหล่งข้อมูลนั้นจะอนุญาตให้โดเมนของ Browsers สามารถเข้าถึงข้อมูลเหล่านั้นได้ (TAeng Trirong, 2017)

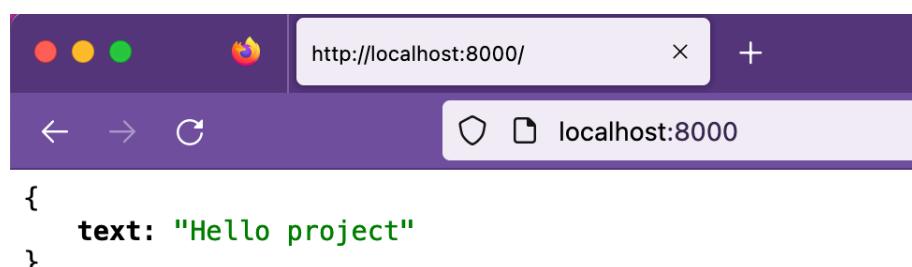
2.1.11 ទடughvī Fastapi

เฟรมเวิร์คสำหรับพัฒนาส่วนต่อประสานเครื่องแม่ข่ายกับเครื่องลูกข่ายด้วยภาษา Python (Python) เฟรมเวิร์ค fastAPI ถูกออกแบบมาให้ง่ายต่อการพัฒนา และสามารถที่จะสร้าง API ขึ้นมาได้อย่างรวดเร็ว โดยประสิทธิภาพการทำงานนั้นเร็ว fastAPI นั้นรองรับการทำงานแบบ Asynchronous และมี Uvicorn เป็นตัว run server ข้อดีของ การใช้งาน fastAPI คือ (Natakorn, 2021)

- มีความเร็วของการทำงานเทียบเท่า Node.js และ Go
 - รูปแบบการเขียนฟังก์ชันต่าง ๆ เช่นง่ายต่อการศึกษา
 - 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 6 ตัวอย่างโค้ดสำหรับการสร้าง Web API ด้วย fastAPI



ภาพที่ 7 ผลลัพธ์แสดงค่าร่วม Hello project จาก fastAPI

2.1.12 ทฤษฎี Git

Git คือ Version Control ที่ถูกพัฒนาขึ้นมาเพื่อใช้ในกระบวนการพัฒนาซอฟต์แวร์ อธิบายให้เข้าใจโดยง่าย คือ ระบบที่ถูกพัฒนาขึ้นมาเพื่อใช้สำหรับการติดตาม ตรวจสอบ การพัฒนา แก้ไข ซอฟต์แวร์ ซอฟต์แวร์ต่าง ๆ ในขั้นตอนการพัฒนา ที่สามารถตรวจสอบได้ทุก ตัวอักษร ทุกบรรทัด ทุกไฟล์ ที่มีการแก้ไข ใครเป็นคนแก้ไข และแก้ไข ณ วันที่เท่าไหร่

ระบบการทำงานของ Git ไม่ได้อยู่แค่การตรวจสอบการแก้ไขเท่านั้น ยังสามารถรวมการ แก้ไขทั้งหมดเข้าด้วยกันได้อย่างชัมฉลาดเรียกว่า CI (Continuous Integration) และใน ปัจจุบัน Git VCS (Version Control System) มีการควบรวมพีเจอร์ที่ทำให้นักพัฒนาทำงานได้สะดวกมากขึ้น สามารถทำงานได้ตั้งแต่ขั้นตอนการพัฒนาไปจนถึงการ Deploy งานขึ้นใช้งานบน Server เรียกว่า CD (Continuous Deployment) รูปแบบการใช้งานของ Git มีด้วยกัน 2 รูปแบบ คือ (codebee, 2020)

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

2.1.13 ทฤษฎี Node.js

Node.js คือสภาพแวดล้อมการทำงานของภาษา JavaScript นอกเว็บเบราว์เซอร์ที่ ทำงานด้วย V8 engine นั่นหมายถึงสามารถใช้ Node.js ในการพัฒนาแอปพลิเคชันแบบ Command line และพลิกแพน Desktop หรือแม้แต่เว็บเซิร์ฟเวอร์โดยที่ Node.js จะมี APIs ที่สามารถ ใช้สำหรับทำงานกับระบบปฏิบัติการ เช่น การรับค่าและการแสดงผล การอ่านเขียนไฟล์ และการ ทำงานกับเน็ตเวิร์ก เป็นต้น

Node.js ถูกพัฒนาและทำงานด้วย Chrome V8 engine สำหรับคอมโพล์ภาษา JavaScript ให้เป็นภาษาเครื่องด้วยการคอมโพล์แบบ Just-in-time (JIT) เพื่อเพิ่มประสิทธิภาพการทำงานของภาษา JavaScript จากที่แต่เดิมมันเป็นภาษาที่มีการทำงานแบบ Interpreted Node.js เป็น โปรแกรมที่สามารถใช้ได้ทั้งบน Windows, Linux และ Mac OS X นั่นหมายความว่าสามารถเขียน โปรแกรมในภาษา JavaScript และนำไปรันได้ทุกระบบปฏิบัติการที่สนับสนุนโดย Node.js นี้ เป็น แนวคิดของการเขียนครั้งเดียวแต่ทำงานได้ทุกที่ (Write once, run anywhere) ข้อดีอีกอย่างหนึ่งใน การใช้ภาษา JavaScript ของ Node.js คือทำให้การพัฒนาเว็บไซต์ทำได้ง่ายขึ้นสำหรับนักพัฒนา เนื่องจากสามารถใช้ภาษา JavaScript สำหรับทั้ง Front-end และ Back-end โดยไม่ต้องศึกษา ภาษาเฉพาะในแต่ละด้าน ตัวอย่างของการพัฒนาเว็บไซต์ในรูปแบบนี้ เช่น React.js ซึ่งเป็นไลบรารี โดย Facebook (ทำความรู้จักกับ Node.js, 2021)

2.1.14 ทฤษฎี Matplotlibs

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

2.1.15 ទடែន MongoDB

MongoDB เป็น open-source document database โดยเป็นฐานข้อมูลแบบ NoSQL คือไม่มีความสัมพันธ์ (No relationship) ของตารางแบบ SQL ทั่วไป แต่จะเก็บข้อมูลเป็นแบบ JSON (JavaScript Object Notation) แทนการบันทึกข้อมูลทุก ๆ Record ใน MongoDB และเรียกว่า Document ซึ่งจะเก็บค่าเป็น key และ value จะเห็นว่าคือ JSON (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 8 ตัวอย่างข้อมูลแบบ JSON

โดยหลัก ๆ เหmagage กับองค์กรที่อยากระเข้าถึงข้อมูลได้อย่างรวดเร็ว อีกทั้งยังเหmagage กับการทำ Big Data และอื่น ๆ ดังนี้ (PLC, 2022)

1. ตัว MongoDB สามารถที่จะสร้างเป็น Cluster เพื่อที่จะตอบสนองของคำว่า High Availability (HA) ได้ ซึ่งสามารถเลือก Region ที่อยากจะ Deploy บน Cloud Provider นั้น ๆ ได้
 2. ความรวดเร็วในการเข้าถึงข้อมูล เพราะว่าตัว Database ของมันไม่มี Schema ซึ่งจะต่างกับ SQL โดยแบบนั้นจะອิงจากฐานข้อมูลที่มาจากการ Table
 3. สามารถทำ Auto Scale ได้ไม่ว่าจะมีการใช้งานมากน้อยแค่ไหน ก็สามารถปรับใช้กับ Environment นั้น ๆ ได้
 4. รองรับการทำ Multiple Cloud Provider ซึ่งข้อดีข้อนี้สามารถทำให้ Database ที่ใช้นั้นมี High Availability มากขึ้นโดยเราไม่จำเป็นที่จะต้องยึดติดกับ Cloud Provider เจ้าใดเจ้าหนึ่ง

2.1.16 ທັງໝົດ Next.js

Next.js เป็น React Web Framework คล้าย ๆ กับ Create React App ที่ช่วยให้เขียนเว็บไซต์ได้สะดวกขึ้น เพราะ Setup และ Config ให้เรียบร้อยครบถ้วน 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 คือ ช่วยในเรื่อง SEO หรือ search engine optimization เพราะถ้าทำการ inspect เว็บไซต์ที่สร้างโดย Next.js จะเห็นว่า source จะเป็น html ส่วนใหญ่ซึ่งทำให้ SEO คนผ่าน source เพื่อให้ได้ข้อมูลและจัดหมวดหมู่ได้ง่ายกว่า React ที่เป็น JavaScript มากกว่า ทำให้ Next.js เป็นที่นิยมในหลาย ๆ บริษัท นอกจากนี้ ข้อดีก็คือ render ได้เร็วกว่า React เพราะ Next.js มีสิ่งที่เรียกว่า get static path ซึ่งการสร้าง path แบบ static แบบเว็บไซต์ html โดยไม่ต้องทำการเชื่อมต่อกับ backend เพื่อให้ได้ data ยังไงก็ได้ สำหรับ Next.js สามารถรวมเข้ากับ backend ได้ง่าย ๆ 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 ทำให้มีไปเพิ่ม size ของ client side (frevation, 2021)

2.1.17 ทฤษฎี Numpy

Numpy เป็น ไลบรารี (Library) ที่รู้จัก และเป็นที่นิยมใช้ในการคำนวณ เช่น ใช้คำนวณ Matrix หรือ คำนวณกับ Array ในงาน Data Science, Data analytics และในการเรียนรู้ของเครื่องจักร (Machine Learning) หรือ ดีป เลิร์นนิ่ง (Deep Learning) ก็ต้องใช้ Numpy ด้วยที่ Numpy เป็น Library ที่มีความสามารถทางคณิตศาสตร์ด้วยภาษา Python สามารถคำนวณ หรือ ดำเนินการทางตรรกศาสตร์ ใน Array หลายมิติหรือ Matrix ได้อย่างรวดเร็ว เพราะ Library เชียนด้วยภาษา C ที่ Compile ไว้แล้ว (mindphp, Numpy คืออะไร)

Numpy นั้นได้แรงบันดาลใจมาจาก MATLAB ดังนั้นผู้ที่มีประสบการณ์ด้าน MATLAB อยู่แล้วจะทำความเข้าใจ Numpy ได้ไม่ยาก โดยหลักการคือการนิยามตัวแปร Array หลายมิติที่คุ้นเคย ในคณิตศาสตร์ อาทิ เช่น เวกเตอร์ (1 มิติ) เมตริก (2 มิติ) เทนเซอร์ (3 มิติขึ้นไป) เป็นต้น และ operations ของมัน ในการทำความเข้าใจ Numpy นั้นควรมีความรู้พื้นฐาน Linear algebra ในเรื่องของ 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 ชุดข้อมูลหลายมิติ (Panel Data) มีจุดเด่นด้านการวิเคราะห์ข้อมูล (Data Analysis) และการทำความสะอาด (Data Cleaning) ซึ่งเป็น Process ที่สำคัญมากในการทำงาน กับข้อมูล Pandas มีความสามารถในการจัดการ และวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพตั้งแต่ ข้อมูลขนาดเล็กไปจนถึงข้อมูลขนาดใหญ่ทำให้ Pandas ตอบโจทย์งานในยุคที่ข้อมูลมีขนาดใหญ่มาก ขึ้นเรื่อยๆ ได้ไม่ปัญหาติดขัดเหมือนกับ Spreadsheets อื่น ๆ เช่น Excel หรือ Google Sheets ซึ่งจะทำงานได้ช้าลงหากข้อมูลมีขนาดใหญ่ขึ้น ขั้นตอนการเตรียมข้อมูลนั้นมีความสำคัญมาก และ

Data Scientist อาจจะใช้เวลาส่วนใหญ่หมดไปกับขั้นตอนนี้ เพราะหากข้อมูลที่เตรียมได้ไม่มีประสิทธิภาพการนำ Insights ไปใช้งานหรือนำข้อมูลไปสร้างโมเดลย่อมทำให้ได้ข้อมูลที่ไม่น่าเชื่อถือ

นอกจากนี้ เมื่อเปรียบเทียบกับ Tools วิเคราะห์ข้อมูลอื่น ๆ อย่าง Excel หรือ Google Sheets อาจไม่ตอบโจทย์เต็มที่หากต้องการเชื่อมต่อกับแหล่งข้อมูลบางประเภท หรือระบบจัดการข้อมูล (Automation) ในขณะที่ pandas ซึ่งเป็นส่วนหนึ่งของ Python นั้นสามารถใช้การเขียนโค้ดเพื่อปรับแต่ง หรือเชื่อมต่อกับโปรแกรมอื่น ๆ ได้สะดวก (Panchart, 2021)

2.1.19 ทฤษฎี Pythainlp

โมดูล pythainlp เป็นเหมือนกับตัว library ที่รวมคำสั่งเกี่ยวกับที่เกี่ยวกับภาษาไทยใน Python ซึ่งก็เป็นตัวช่วยให้การทำงานเกี่ยวกับตัวของภาษาไทย มีประสิทธิภาพและสะดวกมากขึ้น ในการทำงานของ pythainlp ก็จะมีการทำงาน เช่น การตัดคำ การแปลงไทยเป็นอังกฤษ และการเข้าถึงรหัส Soundex และยังมีการทำงานที่เกี่ยวกับตัวของภาษาไทย มีการแสดงผล เกี่ยวกับเซ็ตของภาษาไทย ทั้งหมด ยังมีในส่วนของการแยกด้วย เช่น เป็นส่วนของพยัญชนะ สระ วรรณยุกต์ เป็นต้น ยังมีในส่วนของเลขไทย มีการเช็คว่าเป็นคำภาษาไทยหรือไม่ มีการนับตัวอักษร ว่าเป็นภาษาไทยกี่เปอร์เซ็นต์ และยังมีส่วนของการแสดงคำอ่านที่เป็นพวกรเวลาและยังมีการจัดเรียงคำใน List ให้เรียงกันเป็นลำดับได้ เป็นต้น (mindphp, 2022) นอกจากนี้ยังใช้สำหรับประมวลผลข้อความ และการวิเคราะห์ทางภาษาศาสตร์ NLTK และใช้กับภาษาไทยโดยเฉพาะ มีฟังก์ชันการทำงานที่หลากหลาย เช่น Character Set อักษรไทย คำไทย, เรียงคำภาษาไทย, Stop Words ภาษาไทย, ตัดคำภาษาไทย, วิเคราะห์ชนิดของคำทางไวยากรณ์, ตรวจตัวสะกด แก้คำผิด และอีกมากมาย (Surapong, 2020)

2.1.20 ทฤษฎี Scikit-learn

Scikit-learn หรือ sklearn นำเสนอแบบจำลองทางสถิติและการเรียนรู้ของเครื่องที่หลากหลายแตกต่างจากโมดูลส่วนใหญ่ sklearn ได้รับการพัฒนาใน Python มากกว่า C แม้จะได้รับการพัฒนาใน Python ก็ตาม ประสิทธิภาพของ sklearn นั้นถูกกำหนดให้ใช้ NumPy สำหรับการดำเนินการพื้นฐานเชิงเล็กและอาร์เรย์ที่มีประสิทธิภาพสูง (เจร์, 2021)

Scikit-Learn ถูกสร้างขึ้นโดยเป็นส่วนหนึ่งของโครงการ Summer of Code ของ Google และทำให้ชีวิตของนักวิทยาศาสตร์ข้อมูลที่มี Python เป็นศูนย์กลางนับล้านทั่วโลกง่ายขึ้น ส่วนนี้ของชีรีส์มุ่งเน้นไปที่การนำเสนอด้วยบริการและมุ่งเน้นไปที่องค์ประกอบเดียว นั่นคือการแปลงชุดข้อมูล ซึ่งเป็นขั้นตอนสำคัญและสำคัญที่ต้องทำก่อนพัฒนาแบบจำลองการทำงาน Scikit-learn เป็นแพ็คเกจ Python โอเพนซอร์สพร้อมการวิเคราะห์ข้อมูลที่ซับซ้อน และมาร์กอฟกับอัลกอริทึมในตัวมากมายที่จะช่วยให้คุณได้รับประโยชน์สูงสุดจากโครงการวิทยาศาสตร์ข้อมูลของคุณโดยรา里的 Scikit-learn มีให้เลือกใช้ได้รายการดังนี้ 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 ต่าง ๆ และสามารถทำ Serverless Functions บน Cloud รวมทั้งยังสามารถ Integrate และสร้าง Workflow ผ่าน GitHub เพื่อทำ Automated Deployment โดยได้อ่าย่างง่าย Vercel Inc. เดิมชื่อ Zeit เป็นแพลตฟอร์มคลาวด์ของอเมริกาในฐานะบริษัทผู้ให้บริการบริษัทรักษาการของการพัฒนาเว็บไซต์ Next.js สถาปัตยกรรมของ Vercel สร้างขึ้นจาก Jamstack และการจัดการการปรับใช้ผ่านที่เก็บ Git Vercel เป็นสมาชิกของ MACH Alliance (Huangsri, 2021)

2.2 งานวิจัยที่เกี่ยวข้อง

จักรินทร์ สันติรัตนภักดี และศุภกฤติ์ นิวัฒนาภูล ศึกษาเรื่อง การออกแบบและพัฒนากระบวนการจำแนกชั้นของเรียนรถโดยสารสาธารณะเพื่อติดแท็กปัญหาการให้บริการ องค์การขนส่งมวลชนกรุงเทพ (ขสมก.) มีช่องทางในการร้องเรียนรถโดยสารสาธารณะผ่านเว็บบอร์ด ที่ผู้ใช้งานสามารถแสดงความคิดเห็นได้อย่างอิสระ ผู้วิจัยจึงออกแบบและพัฒนากระบวนการจำแนกชั้นของเรียนรถโดยสารสาธารณะ จากชั้นร้องเรียนผ่านเว็บบอร์ดขององค์การขนส่งมวลชนกรุงเทพด้วยกระบวนการตัดคำภาษาไทยโดยใช้พจนานุกรม แล้วคัดเลือกคำที่ตัวการวิเคราะห์นำหัวหักของคำมาสร้างเป็นคลังคำที่แบ่งเป็น 4 คลาส ได้แก่ คลาสการขับขี่ คลาสผู้ขับขี่ และพนักงานผู้ให้บริการ คลาสยกพานะและอุปกรณ์ให้บริการ และคลาสเวลาและการเดินรถโดยใช้มอดูลการตัดคำภาษาไทย (Thai Word Segmentation) ด้วยชั้นความทั่วไปซึ่งอยู่ในรูปแบบประโยคมาแบ่งออกเป็นคำหรือคุณลักษณะ (Term/Feature) เพื่อแยกส่วนของชั้นความของจากกันก่อนนำไปประมวลผลในขั้นต่อไป แบ่งตามกระบวนการทำงานออกเป็น 3 กลุ่ม ได้แก่ 1) การตัดคำโดยใช้กฎ (Rule-Based Approach) 2) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) 3) การตัดคำโดยใช้คลังคำที่ (Corpus-Based Approach) จากการทดลองพบว่าอัลกอริทึมโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น มีค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวมสูงที่สุด (จักรินทร์ สันติรัตนภักดี, 2021)

กุณิชัย วิเชียรไชย ศึกษาเรื่อง การเปรียบเทียบวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างการเขียนกับโครงสร้างพยานค์ งานวิจัยนี้นำเสนอการแบ่งแยกคำภาษาไทยโดยเทียบกับโครงสร้างการเขียนของภาษาไทยและอัลกอริทึมการแบ่งแยกคำภาษาไทยโดยโครงสร้างพยานค์ เพื่อศึกษาและเปรียบเทียบวิธีการประมวลผลของการแบ่งแยกคำภาษาไทย และประสิทธิภาพความถูกต้องของอัลกอริทึม โดยสามารถแบ่งงานวิจัยในการแบ่งแยกคำภาษาไทยได้เป็นดังนี้ คือวิธีการใช้กฎ (Rule base approach) วิธีการใช้อัลกอริทึม (Algorithm ap-proach) วิธีการใช้พจนานุกรม (Dictionary base approach) และวิธีการใช้คลังข้อความ (Corpus based approach) ผู้วิจัยจึงได้เสนอวิธีการแบ่งแยกคำภาษาไทย โดยใช้โครงสร้างการเขียนภาษาไทยเพื่อแก้ไขลดพื้นที่ในการจัดเก็บคำที่ในพจนานุกรม และวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างพยานค์เพื่อลดการสิ้นเปลืองพื้นที่ในการจัดเก็บพจนานุกรม ยกตัวอย่างการแบ่งแยกคำและพยานค์ของคำว่า “ประเทศไทย” จะสามารถ

แบ่งแยกคำได้เป็น “ประตูเทคโนโลยี” และแบ่งพยางค์ได้เป็น “ประตูเทคโนโลยี” จากผลลัพธ์ในการแบ่งแยกคำนั้นยังขาดความถูกต้องในการแบ่งแยกคำซึ่งสามารถพัฒนาแนวคิดในการศึกษาและสร้างกฎเพื่อแบ่งแยกคำให้ถูกต้องมากยิ่งขึ้น (วุฒิชัย, 2013)

ปราณี พึงวิชา อานันท์ ทับเที่ยง และธัญญา สัตยาภิธาน (2019) ศึกษาการแบ่งกลุ่มพุทธิกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ ผู้บริโภคด้วยวิธีการจัดกลุ่มด้วยเคลื่อน (K-Means Clustering) เป็น 2 กลุ่ม ซึ่งมีลักษณะเฉพาะในแต่ละกลุ่ม จากการวิเคราะห์ความแตกต่างระหว่างกลุ่มของหัตถศิลป์ด้านพุทธิกรรมการซื้อและด้านส่วนประสมทางการตลาดที่มีผลต่อการตัดสินใจซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์เมื่อยื่นต่างกัน โดยการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA) พบว่า ด้านพุทธิกรรมการซื้อทั้ง 2 กลุ่ม มีความถี่ในการซื้อต่างๆ เช่นไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนตัวแปรอื่น ๆ นั้นมีความแตกต่างกัน จากการวิเคราะห์แบ่งกลุ่มผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ด้วยวิธี K-mean clustering สามารถจำแนกเป็น 2 กลุ่มโดยแต่ละกลุ่มนี้มีลักษณะเฉพาะดังนี้ กลุ่มที่ 1 : กลุ่มกระเปาหนักจ่ายได้ถ้าชอบไม่ค่อยชอบออกสื่อ ลักษณะด้านประชากรศาสตร์ โดยส่วนใหญ่เป็นคน Generation X เพศหญิงมากกว่าเพศชาย มีระดับการศึกษาสูงกว่าบริษัทฯ กลุ่มที่ 2 : กลุ่มวัยสะอ่อน ชอบออกสื่อ ซื้อน้อยแต่บ่อยครั้ง ลักษณะด้านประชากรศาสตร์ โดยส่วนใหญ่เป็นคน Generation Y เพศหญิงมากกว่าเพศชาย ส่วนใหญ่มีระดับการศึกษาสูงแต่จะน้อยกว่ากลุ่ม 1 โดยมีระดับปริญญาตรีมากที่สุดโดยส่วนใหญ่เป็นพนักงานบริษัทเอกชน (ปราณี พึงวิชา, 2019)

ธงชัย คล้ายคลึง วุฒิชัย สง่างาม กิตติวงศ์ สุธรรมโน และพันธุ์พงศ์อภิชาตกุลศ (2019) ศึกษาเรื่อง เทคนิคการคัดเลือกกลุ่ม โหลดรายอาชารสำหรับรับแผนการติดตั้งระบบผลิตไฟฟ้า พลังงานแสงอาทิตย์ บนหลังคาเพื่อเพิ่มค่าครองใช้พลังงานไฟฟ้า บทความนี้ต้องการนำเสนอเทคนิควิธีการคัดเลือกกลุ่มโหลดในแต่ละอาคารที่มีความเหมาะสมสำหรับติดตั้งระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์กรณีที่ติดตั้งบนหลังคาของอาคารในศูนย์กลางมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ในการวิเคราะห์การ群聚ของ K-Means Clustering เริ่มต้นด้วยการจัดแบ่งข้อมูลออกเป็น K กลุ่ม กำหนดจุดศูนย์กลางเริ่มต้นจำนวน K จุดขึ้นตอนต่อไปคือการสร้างกลุ่มข้อมูลและความสัมพันธ์กับจุดศูนย์กลางที่ใกล้มากที่สุด จากผลการวิเคราะห์การใช้พลังงานไฟฟ้าทั้งหมด 34 อาคารด้วยวิธีการ K-Mean Clustering ทำให้สามารถแยกแยะจัดกลุ่มโหลดรายอาคารได้อย่างมีประสิทธิภาพซึ่งแบ่งได้ 3 กลุ่มโดยกลุ่มที่ 3 จำนวน 19 อาคารนั้นเป็นกลุ่มอาคารที่มีความเหมาะสมทั้งด้านพุทธิกรรมการใช้พลังงานไฟฟ้า และมีพื้นที่รองรับการติดตั้งระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์บนหลังคาได้ (Thongchai Klayklueng, 2019)

วรกร ประดิษฐ์กุล ปราลี มณีรัตน์ และ นิเวศ จิระวิชิตชัย (2021) ศึกษาเรื่อง ระบบแนะนำรายนต์ให้กับลูกค้าโดยการวิเคราะห์จากการอ้างอิงถึงพุทธิกรรมของผู้ใช้ (Collaborative Filtering) กรณีศึกษาบริษัท โตโยต้า บัสส์ จำกัด ผู้วิจัยได้พัฒนาระบบแนะนำรายนต์ให้กับลูกค้าโดยการวิเคราะห์จากการอ้างอิงถึงพุทธิกรรมของผู้ใช้ เพื่อช่วยให้ลูกค้าได้รับการแนะนำรุ่นรถยนต์ที่

เหมาะสม ตรงตามความต้องการของลูกค้า ผู้วิจัยใช้อัลกอริทึมการหาความคล้ายคลึงกันของผู้ใช้ โดยวิเคราะห์จากลูกค้าที่มีพฤติกรรมใกล้เคียงกันด้วยสมการการหาความคล้ายโคไซน์ ซึ่งเป็นฟังก์ชันในภาษา Python ในการพัฒนาระบบแนะนำรายนี้ให้กับลูกค้า ด้วยสมการความคล้ายโคไซน์ (cosine similarity) จากการทดลองเมื่อนำข้อมูลมาจัดลำดับคะแนนความชอบของผู้ใช้แต่ละคน เพื่อเป็นการเพิ่มความเร็วให้อัลกอริทึมของวิธีการกรองแบบร่วมมือ อีกทั้งระบบจะนำค่าความคล้ายคลึง โคไซน์ของผู้ใช้ในระบบกับผู้ใช้เป้าหมายมาทดสอบความแม่นยำของระบบด้วยค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ พบร่วมกัน 0.97 เมื่อกำหนดค่า k ไว้เท่ากับ 5 สรุปได้ว่าระบบมีประสิทธิภาพในการแนะนำร้านเดียวกันและความแม่นยำอยู่ในระดับที่ดี (Warakorn Pradiskul, 2021)

บทที่ 3

ขั้นตอนการดำเนินงาน

การจัดทำโครงการทางวิทยาการคอมพิวเตอร์ ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลมีน (K-Means) ผู้จัดทำได้กำหนดขั้นตอนการดำเนินงานดังนี้

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

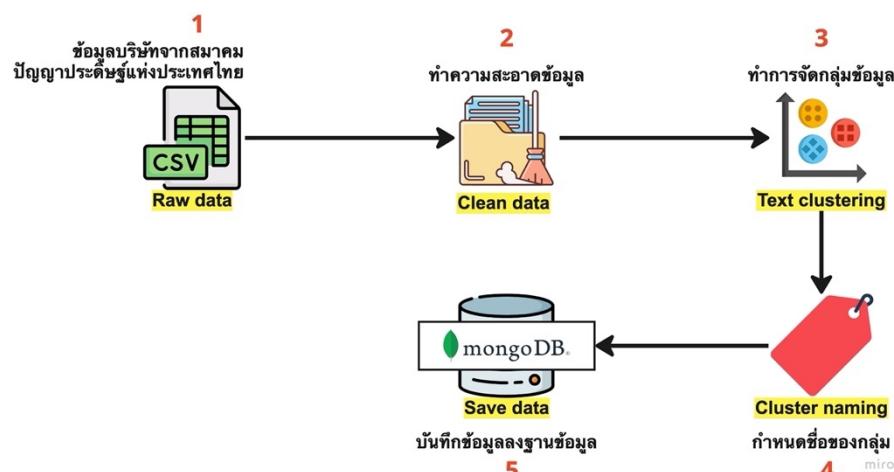
3.1 การเตรียมและวิเคราะห์ข้อมูล

ตาราง 4 การวิเคราะห์ข้อมูล

ข้อมูล	จำนวน	หน่วย
จำนวนข้อมูลสถานประกอบการทั้งหมด	1,643	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

ตาราง 4 (ต่อ)

ข้อมูล	จำนวน	หน่วย
biometrics	9	รายการ
biometric authentication	3	รายการ
sentiment analysis	3	รายการ
text mining	2	รายการ
embedded system	1	รายการ
machine translation	1	รายการ
ไม่มีประเภท	1,318	รายการ
จำนวนคำทั้งหมด	9.856	คำ



ภาพที่ 9 การเตรียมและวิเคราะห์ข้อมูล

จากการที่ 9 แสดงการเตรียมและวิเคราะห์ข้อมูลได้ดังนี้

1. ข้อมูลบริษัทจากสมาคมปัญญาประดิษฐ์แห่งประเทศไทยโดยเป็นไฟล์ข้อมูลแบบ CSV (Comma-Separated Value)
2. ทำความสะอาดข้อมูลลบข้อมูลที่ไม่มีความหมายในตัว ลบตัวเลขที่ไม่จำเป็น แก้ไขพิมพ์ผิดและอักษรพิเศษต่าง ๆ
3. จัดกลุ่มข้อมูลด้วยวิธี K-Means clustering
4. ตั้งชื่อของกลุ่มข้อมูลโดยอ้างอิงจากประเภทงานด้านใดๆ จากเว็บไซต์ th.jobsdb.com
5. นำผลการจัดกลุ่มจัดเก็บลงฐานข้อมูล MongoDB

3.1.1 การนำเข้าไฟล์ .csv เพื่อทำการจัดกลุ่มข้อมูล

1. ดาวน์โหลดโปรเจคจาก https://github.com/slapexs/final_project
2. 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 10 การเรียกใช้ไลบรารี (Library) สำหรับคำนวณค่า TF-IDF

จากภาพที่ 10 แสดงการเรียกใช้พัฟฟ์ชันจากไลบรารี (Library) ที่ใช้ในการคำนวณค่า TF-IDF ประกอบไปด้วย

1. Pandas ใช้ในการอ่านข้อมูลในไฟล์
 2. thai_stopword เป็นรายการคำที่ไม่สื่อความหมายในภาษาไทย
 3. stopword เป็นรายการคำที่ไม่สื่อความหมายในภาษาอังกฤษ
 4. word_tokenize เป็นฟังก์ชันที่ใช้ในการตัดคำแยกเป็นคำ ๆ จากประโยค
 5. TfidfVectorizer เป็นฟังก์ชันสำหรับคำนวณหาค่า TF-IDF จากประโยคที่ตัดคำแล้ว
 6. String เป็นคลาสของภาษาไพธอนที่ใช้แสดงข้อมูลตัวอักษรต่าง ๆ
 7. Numpy ใช้ในการสร้างอาร์เรย์สำหรับการใช้งานในการอ่านข้อมูล

ภาพที่ 11 การอ่านข้อมูลจากไฟล์และกำหนดตัวกรองการตัดคำ

จากภาพที่ 11 แสดงการอ่านข้อมูลจากไฟล์ และการกำหนดตัวกรอกในการตัดคำทั้งคำที่ไม่สื่อความหมายในภาษาไทย ภาษาอังกฤษ ตัวเลขไทย และอักขระพิเศษ พร้อมทั้งประกาศตัวแปรเพื่อเก็บค่าการจัดการตัดคำ

```

● ● ●
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 12 พัฒนาสำหรับใช้ลบตัวเลข และอักขระพิเศษ

จากภาพที่ 12 แสดงการลบตัวเลข และอักขระพิเศษออกจากประโภคที่รับเข้ามาและทำการเชื่อมประโภคและคืนค่ากลับออกไป

```

● ● ●
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 13 พัฒนาสำหรับใช้ลบคำที่ไม่สื่อความหมายและตัวเลขไทย

จากภาพที่ 13 แสดงการลบคำที่ไม่สื่อความหมายในภาษาไทย ภาษาบัญถงกฤษ 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 14 การวนซ้ำข้อมูลเพื่อตัดคำและทำความสะอาดข้อมูล

จากภาพที่ 14 แสดงการวนซ้ำการส่งข้อมูลที่อ่านจากไฟล์เพื่อนำไปลบคำที่ไม่สื่อความหมาย ตัวเลข และอักขระพิเศษออกจากข้อมูลตัวแรกถึงตัวสุดท้าย 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 15 การเทรนและการทดสอบโมเดลการคำนวณค่า TF-IDF

จากภาพที่ 15 แสดงการสร้างเวกเตอร์ของการคำนวณค่า TF-IDF การเทรนข้อมูล และการทดสอบการประมวลผลจากโมเดลที่เทรนประกอบไปด้วย

1. vectorizer เป็นการเทรนโมเดลสำหรับการคำนวณค่า TF-IDF
2. tfidf_vector เป็นการทดสอบและสร้างเวกเตอร์ของคำในแต่ละประโยค
3. tfidf_array เป็นการนำเวกเตอร์มาเปลี่ยนให้อยู่ในรูปแบบของอาเรย์
4. df_tfidf เป็นการนำข้อมูลในอาเรย์มาสร้างเป็นตารางข้อมูล

ภาพที่ 16 ผลลัพธ์การคำนวณค่า TF-IDF

จากภาพที่ 16 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 17 การเรียกใช้ไลบรารีสำหรับการ จัดกลุ่มข้อมูลด้วยเคมีน (K-Means)

จากภาพที่ 17 แสดงการนำเข้าไลบรารีจาก Scikit-learn เพื่อทำการจัดกลุ่มข้อมูลโดยมีรายละเอียดดังนี้

1. KMeans เป็นการเรียกใช้ลักษณะที่มีสำหรับการจัดกลุ่ม
2. PCA (Principle Components Analysis) ซึ่งเป็นวิธีการลด dimension ของ Feature ลงช่วยลดTHONความซ้ำซ้อนของข้อมูลทำให้ 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 18 แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k และการจัดกลุ่มข้อมูล

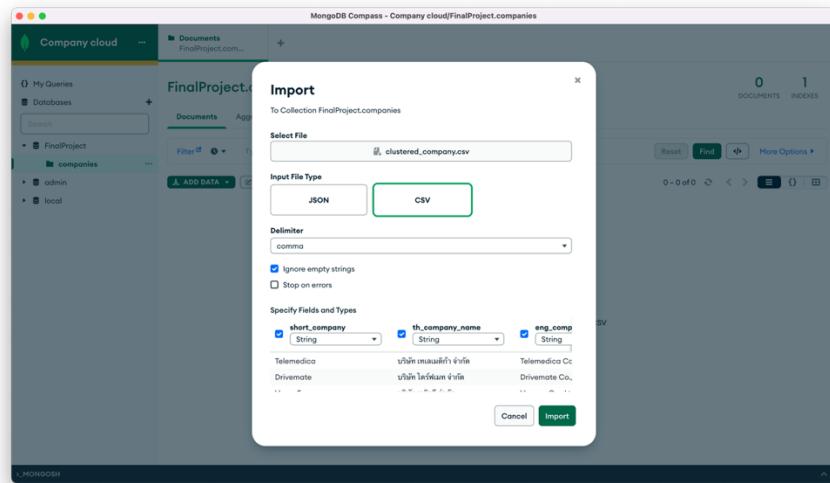
4. เมื่อกำหนดค่าการจัดกลุ่มเรียบร้อยเรียกใช้ไฟล์ clustering.py ใน Terminal เพื่อทำการจัดกลุ่มและบันทึกผลลัพธ์

```
onze@Tinngrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh) ⌘1 ..project/report (-zsh) ⌘2 +
15% 4.4 GB main
~/desktop/final_project main ↵ backup_finalproject 11:37:48
python clustering.py
```

ภาพที่ 19 การเรียกใช้งานไฟล์ clustering.py เพื่อจัดกลุ่มข้อมูลและบันทึกผลลัพธ์

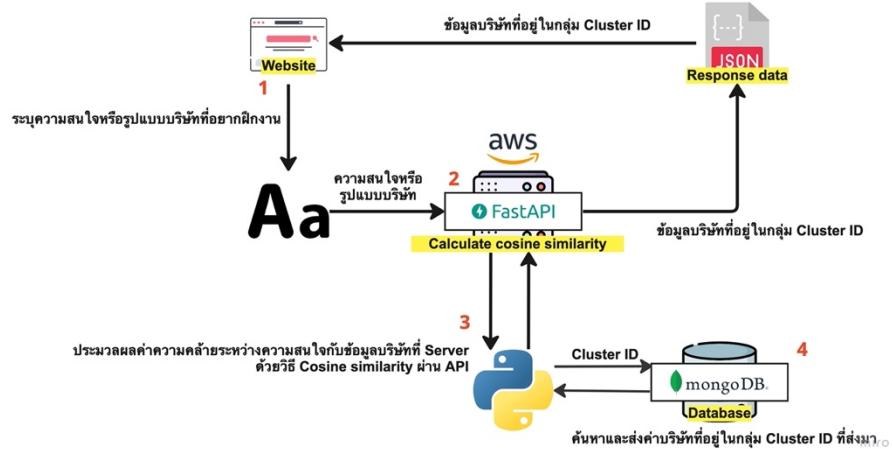
3.1.4 การนำเข้าข้อมูล Clustering เข้าสู่ฐานข้อมูล

1. นำเข้าข้อมูลลงฐานข้อมูล MongoDB โดยโปรแกรม MongoDB Compass
2. เชื่อมต่อ MongoDB Compass กับ Mongodb Atlas
3. นำเข้าข้อมูลด้วยไฟล์ .CSV ที่เป็นผลลัพธ์จากการจัดกลุ่มข้อมูล



ภาพที่ 20 การนำเข้าข้อมูลลงสู่ฐานข้อมูล MongoDB

3.2 การทำงานของระบบ



ภาพที่ 21 การทำงานของระบบ

จากภาพที่ 21 แสดงการทำงานของระบบได้ดังนี้

1. Website ใช้ระบุความสนใจเพื่อส่งค่าไปประมวลผลความคล้ายคลึงกับข้อมูลบริษัท
2. Server ใช้ประมวลผลความคล้ายคลึงกันระหว่างความสนใจที่ได้รับมา และข้อมูลบริษัทที่อยู่ในฐานข้อมูลด้วยเทคนิค Cosine similarity โดยภาษา Python และส่งค่ากลับไปเป็น Cluster ID
3. เมื่อได้ Cluster ID แล้วนำไปค้นหาบริษัทที่ Cluster ID ตรงกันในฐานข้อมูลและคืนค่า Response API เป็นข้อมูลในรูปแบบ JSON ที่มีข้อมูลบริษัทที่อยู่ใน Cluster ID นั้น
4. MongoDB เป็นฐานข้อมูลที่เก็บข้อมูลบริษัทไว้ และรอให้เซิร์ฟเวอร์เรียกใช้ข้อมูลเพื่อนำไปแสดงผล

3.3 การวิเคราะห์และออกแบบระบบ

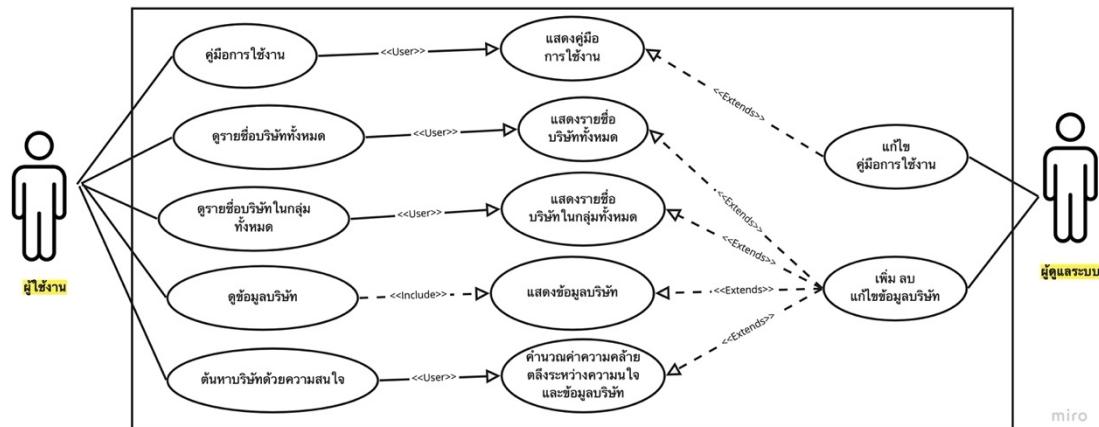
3.3.1 การวิเคราะห์ระบบ

การวิเคราะห์ระบบและการออกแบบ (System Analysis and Design) คือ วิธีการที่ใช้ในการสร้างระบบสารสนเทศขึ้นมาใหม่ในธุรกิจ ได้แก่ วิธีการที่ใช้ในการสร้างระบบสารสนเทศใหม่แล้ว การวิเคราะห์ระบบ ช่วยในการแก้ไขระบบสารสนเทศเดิมที่มีอยู่แล้วให้ดีขึ้น การวิเคราะห์ระบบ คือ การหาความต้องการ (Requirements) ของระบบสารสนเทศว่า คืออะไร หรือต้องการเพิ่มเติมอะไรเข้ามาในระบบ การออกแบบ คือ การนำเอาความต้องการของระบบมาเป็นแบบแผน หรือเรียกว่าพิมพ์เขียวในการสร้างระบบสารสนเทศนั้นให้ใช้งานได้จริง

3.3.2 ผู้ใช้และโครงสร้าง (Use Case Diagram)

ผู้ใช้และโครงสร้าง (Use Case Diagram) คือ แผนภาพที่ใช้แสดงปฏิสัมพันธ์ระหว่างระบบงานและสิ่งที่อยู่นอกระบบงาน และแสดงให้เห็นถึงส่วนประกอบทั้งหมด หรือ ภาพรวมของระบบ เป็นรากฐานในการเริ่มต้นการวิเคราะห์ระบบ โดยคุณภาพที่สำคัญที่สุดคือ ความชัดเจนและถูกต้อง

การทำงานหรือเทคนิคการทำงานเปรียบเสมือน “กล่องดำ” โดย Use Case Diagram จะช่วยให้ผู้พัฒนาระบบสามารถแยกแยะกิจกรรมที่อาจจะเกิดขึ้นในระบบ เป็น Diagram พื้นฐาน ที่สามารถอธิบายสิ่งต่าง ๆ ได้โดยใช้รูปภาพที่ไม่ซับซ้อน



ภาพที่ 22 Use Case Diagram ของระบบ

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

ตาราง 6 (ต่อ)

Purpose:	ดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Pre-condition:	เมื่อต้องการดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Main flow:	<ol style="list-style-type: none"> ผู้ใช้งานเปิดเว็บไซต์เข้าไปยังหน้ารายชื่อบริษัททั้งหมด ดูรายชื่อบริษัททั้งหมดที่มีในฐานข้อมูล
Alternate/Exceptional Flow:	None

ตาราง 7 คำอธิบาย Use case ดูรายชื่อบริษัทในกลุ่มทั้งหมด

Use case id:	3
Use case name:	ดูรายชื่อบริษัทในกลุ่มทั้งหมด
Actor:	ผู้ใช้งาน
Scenario:	การดูรายชื่อบริษัทในกลุ่มทั้งหมด
Trigger event:	กรณีที่แสดงผลจากการค้นหาด้วยความสนใจ หรือกรณีที่คลิกเมนูกลุ่มของบริษัท
Brief Description:	ดูรายชื่อบริษัทในกลุ่มที่ต้องการทั้งหมด
Purpose:	เพื่อดูรายชื่อบริษัทในกลุ่มทั้งหมด
Pre-condition:	เมื่อต้องการดูรายชื่อบริษัทในกลุ่มทั้งหมด
Main flow:	<ol style="list-style-type: none"> ผู้ใช้งานเปิดเว็บไซต์เข้าไปยังหน้ากลุ่มบริษัทที่ต้องการ 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:	<ol style="list-style-type: none"> ผู้ใช้งานคลิกที่เมนูชื่อของบริษัทที่ต้องการดูข้อมูล 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
Purpose:	เพื่อคนหาบริษัทด้วยความสนใจ
Pre-condition:	เมื่อต้องการคนหาบริษัทด้วยความสนใจ
Main flow:	<ol style="list-style-type: none"> 1. ผู้ใช้งานพิมพ์ความสนใจหรือรูปแบบธุรกิจที่ช่องค้นหาและ 2. แสดงรายชื่อบริษัทที่อยู่ในกลุ่มที่ระบบแนะนำ 3. ผู้ใช้คลิกเลือกบริษัทเพื่อดูข้อมูลบริษัท
Alternate/Exceptional Flow:	None

ตาราง 10 คำอธิบาย Use case แก้ไขคุณมีการใช้งาน

Use case id:	6
Use case name:	แก้ไขคุณมีการใช้งาน
Actor:	ผู้ดูแลระบบ
Scenario:	การแก้ไขคุณมีการใช้งาน
Trigger event:	None
Brief Description:	แก้ไขคุณมีการใช้งาน
Purpose:	เพื่อแก้ไขคุณมีการใช้งาน
Pre-condition:	เมื่อต้องการแก้ไขคุณมีการใช้งาน
Main flow:	<ol style="list-style-type: none"> 1. ผู้ดูแลระบบแก้ไขข้อมูลคุณมีการใช้งาน 2. Deploy เพื่ออัปเดตระบบ
Alternate/Exceptional Flow:	None

ตาราง 11 คำอธิบาย Use case เพิ่ม ลบ แก้ไขข้อมูลบริษัท

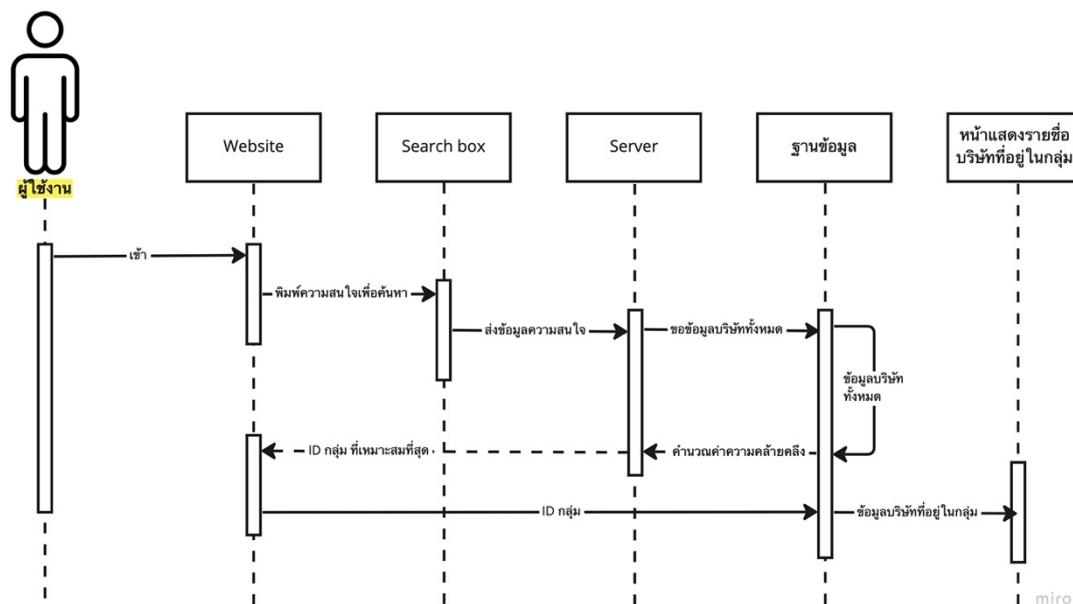
Use case id:	7
Use case name:	เพิ่ม ลบ แก้ไขข้อมูลบริษัท
Actor:	ผู้ดูแลระบบ
Scenario:	การเพิ่ม ลบ แก้ไขข้อมูลบริษัท
Trigger event:	None
Brief Description:	เพิ่ม ลบ แก้ไขข้อมูลบริษัท

ตาราง 11 (ต่อ)

Purpose:	เพื่อเพิ่ม ลบ แก้ไขข้อมูลบริษัท
Pre-condition:	เมื่อต้องการเพิ่ม ลบ แก้ไขข้อมูลบริษัท
Main flow:	1. ผู้ดูแลระบบเพิ่ม ลบ แก้ไขข้อมูลบริษัท ในฐานข้อมูล
Alternate/Exceptional Flow:	None

3.2.3 ซีเคแวนซ์ไดอะแกรม (Sequence Diagram)

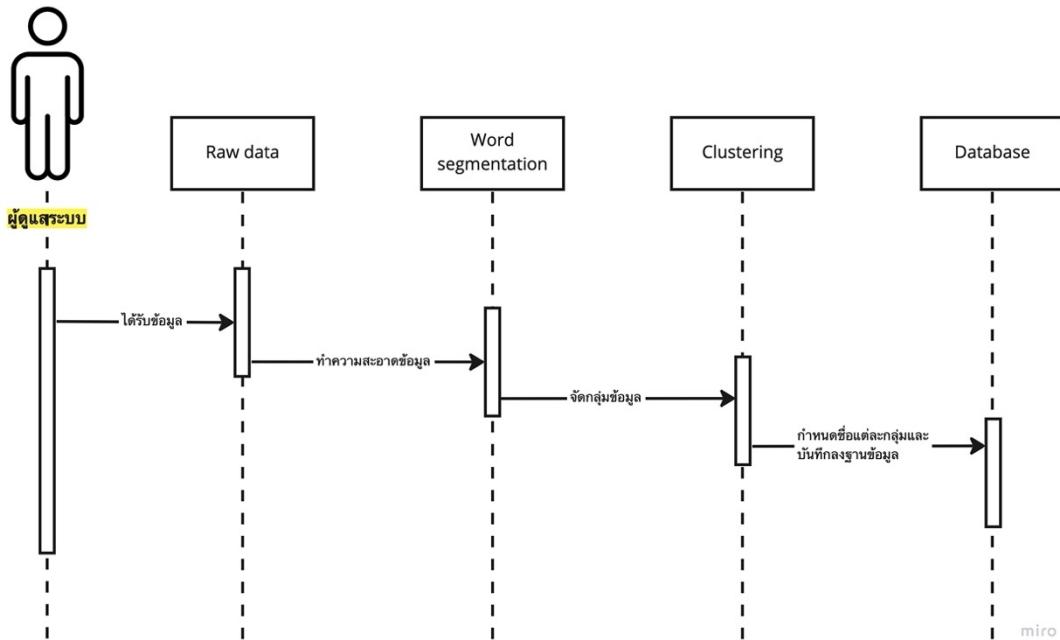
Sequence Diagram เป็นหนึ่งในแผนผังการทำงานแบบ Unified Modeling Language (UML) ใช้สำหรับการสร้างแบบจำลองเชิงวัตถุ โดยขอแตกต่างจากแผนผังรูปแบบ UML อื่นคือเป็น แผนผังการทำงานที่แสดงลำดับการปฏิสัมพันธ์ (Sequence of interactions) ระหว่างวัตถุที่แสดงภายในระบบต่าง ๆ เช่น การส่งข้อความ (messaging) ที่มีการรับส่งข้อมูลระหว่างผู้ใช้ Sequence Diagram เป็นแผนผังการทำงานที่ประกอบไปด้วยคลาส (Class) หรือวัตถุ (Object) เส้นประที่ใช้เพื่อแสดงลำดับเวลา และเส้นที่ใช้เพื่อแสดงกิจกรรมที่เกิดขึ้นจากคลาสหรือวัตถุในแผนผังการทำงานภายใน Sequence Diagram จะใช้สีเหลี่ยมแทนสมือนคลาสและวัตถุโดยภายในจะมีชื่อของคลาส หรือวัตถุประกอบอยู่ในรูปแบบ [Object]: Class



ภาพที่ 23 Sequence Diagram การค้นหาบริษัทด้วยความสนใจของผู้ใช้

ตาราง 12 อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การค้นหาบริษัทด้วยความสนใจ

เหตุการณ์ที่เกิดขึ้น	คำอธิบาย
เข้า Website	เข้า Website ด้วย Browser
พิมพ์ความสนใจเพื่อค้นหา	ระบุความสนใจรูปแบบธุรกิจของบริษัทหรือความสนใจที่อยากฝึกงานของผู้ใช้
ส่งข้อมูลความสนใจ	ส่งข้อมูลความสนใจที่ผู้ใช้ระบุไปประมวลผลที่ Server
ขอข้อมูลบริษัททั้งหมด	Server ขอข้อมูลบริษัททั้งหมดจากฐานข้อมูลเพื่อนำมาเก็บไว้รอคำนวนคาดคะเนคล้ายคลึง
ข้อมูลบริษัททั้งหมด	ข้อมูลบริษัททั้งหมดในฐานข้อมูล ส่งให้ Server
คำนวนคาดคะเนคล้ายคลึงระหว่างความสนใจของผู้ใช้และข้อมูลบริษัท	คำนวนคาดคะเนคล้ายคลึงระหว่างความสนใจของผู้ใช้และข้อมูลบริษัท
ID กลุ่มที่เหมาะสมที่สุด	คืนค่า ID ของกลุ่มบริษัทที่คล้ายกับความสนใจของผู้ใช้
ID กลุ่ม	ส่งค่า ID ของกลุ่มไปยังฐานข้อมูลเพื่อขอข้อมูลบริษัทที่อยู่ในกลุ่มนั้น ๆ
ข้อมูลบริษัทที่อยู่ในกลุ่ม	แสดงรายชื่อบริษัทที่อยู่ในกลุ่มในหน้าเว็บ



ภาพที่ 24 Sequence Diagram การเพิ่มข้อมูลและจัดกลุ่มบริษัทใหม่

ตาราง 13 อธิบายเหตุการณ์ที่เกิดขึ้นใน Sequence Diagram การเพิ่มข้อมูลและจัดลุ่มใหม่

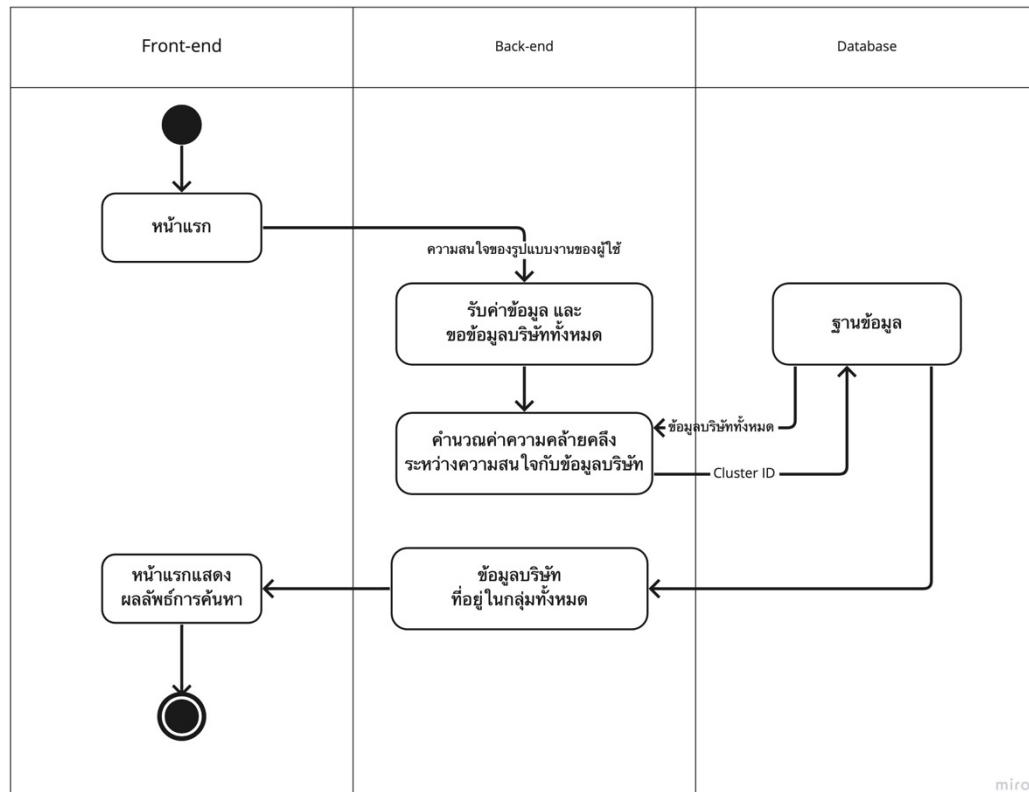
เหตุการณ์ที่เกิดขึ้น	คำอธิบาย
ได้รับข้อมูล	ได้รับข้อมูลดิบที่จะนำมาใช้งาน
ทำความสะอาดข้อมูล	นำข้อมูลดิบมาทำการลบตัวเลข คำที่ไม่มีความหมายในตัว คำສากผิด
จัดลุ่มข้อมูล	ทำการหาคำสำคัญและทำการจัดลุ่มข้อมูล
กำหนดชื่อแต่ละกลุ่มและบันทึกลงฐานข้อมูล	กำหนดชื่อของกลุ่มและบันทึกข้อมูลลงฐานข้อมูลเพื่อใช้ในเว็บไซต์

3.3.4 แอคทิวิตี้ไดอะแกรม (Activity Diagram)

Activity Diagram หรือแผนภาพกิจกรรม ใช้อธิบายกิจกรรมที่เกิดขึ้นในลักษณะกราฟแสดงการไหลของการทำงาน (Workflow) จะมีลักษณะเดียวกับ Flowchart โดย ขั้นตอนในการทำงานแต่ละขั้นจะเรียกว่า Activity

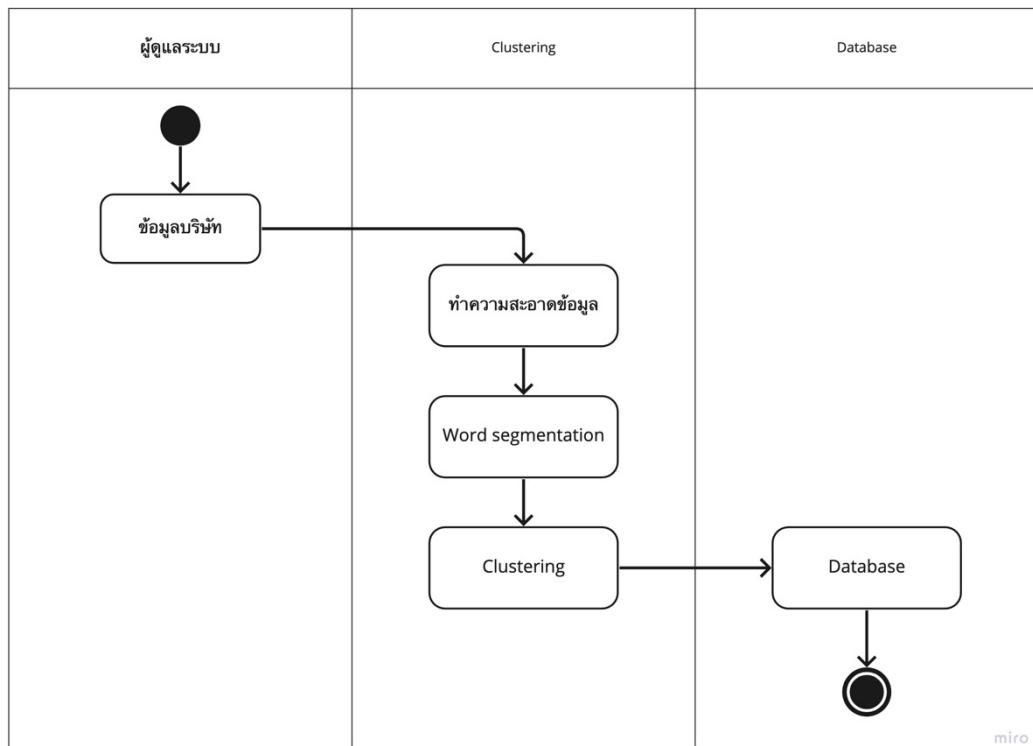
การใช้งาน Activity Diagram

- อธิบายกราฟแสดงการไหลของการทำงาน (Workflow)
- แสดงขั้นตอนการทำงานของระบบ



ภาพที่ 25 Activity Diagram ของผู้ใช้งาน

จากภาพที่ 25 แสดงขั้นตอนการทำงานของระบบเมื่อผู้ใช้คนหนึ่งที่ด้วยความสนใจของผู้ใช้การทำงานจะเริ่มต้นจากการที่ผู้ใช้ระบุความสนใจ จากนั้นทำการส่งข้อมูลไปคำนวณความคล้ายคลึงกัน API และเรียกข้อมูลบริษัทที่มีความคล้ายมากที่สุดมาแสดงผลหน้าเว็บไซต์



ภาพที่ 26 Activity Diagram ของผู้ดูแลระบบ

จากภาพที่ 26 แสดงการทำงานของการจัดกลุ่มข้อมูลเพื่อนำไปใช้ในระบบโดยการทำงานเริ่มต้นที่นำข้อมูลที่ได้เก็บรวบรวมลงในไฟล์จากนั้นทำการทำ Word segmentation และทำการจัดกลุ่มข้อมูล สุดท้ายบันทึกข้อมูลเพื่อนำไปนำเข้าลงฐานข้อมูล MongoDB

3.4 การออกแบบฐานข้อมูล

3.4.1 แบบจำลองโครงสร้างของฐานข้อมูล (Entity–Relationship Diagrams:ER Diagram)

แบบจำลองโครงสร้างของฐานข้อมูล (Entity–Relationship Diagrams:ER Diagram) เป็นเครื่องมือที่ใช้ในการออกแบบฐานข้อมูล ให้สามารถแสดงความสัมพันธ์ของข้อมูล (Relationship) ประกอบด้วย 1. เอกชนิติ (Entity) เป็นวัตถุ หรือสิ่งของที่เราสนใจในระบบงานนั้น 2. แอทริบิวต์ (Attribute) เป็นคุณสมบัติของวัตถุที่เราสนใจ 3. ความสัมพันธ์ (Relationship) คือ ความสัมพันธ์ระหว่างเอกชนิติ ER Diagram มีความสำคัญต่อการพัฒนาระบบงานฐานข้อมูล Application ต่าง ๆ ที่ต้องการการเก็บข้อมูลอย่างมีระบบ มีโครงสร้าง ดังนี้ ER Diagram จึงใช้เพื่อเป็นเอกสารในการสื่อสารระหว่างนักออกแบบระบบ และนักพัฒนาระบบ เพื่อให้สื่อสารอย่างตรงกัน

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

ภาพที่ 27 ER Diagram ระบบแนะนำสถานที่ทำงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยี การจัดการลุ่มเดเมิน (K-Means)

3.4.2 พจนานุกรมข้อมูล (Data Dictionary)

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

ตาราง 14 พจนานุกรมข้อมูลบริษัท

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

ตาราง 14 (ต่อ)

type_innovation	String	50	ตัวอักษร	ประเภท เทคโนโลยี	Null
detail	String	255	ตัวอักษร	รายละเอียด ธุรกิจ	Null
owner	String	255	ตัวอักษร	เจ้าของ	Null
province_base	String	50	ตัวอักษร	จังหวัดที่ตั้ง	Null
address	String	255	ตัวอักษร	ที่อยู่โดย ละเอียด	Null
phone_number	String	20	ตัวอักษร	เบอร์โทร	Null
email	String	50	ตัวอักษร	อีเมล	Null
website	String	50	ตัวอักษร	เว็บไซต์	Null
source	String	255	ตัวอักษร	ที่มาข้อมูล	Null
cluster	String	1	ตัวอักษร	กลุ่ม	Not Null

3.5 การออกแบบหน้าจอ

การออกแบบหน้าจอหรือ UI design นั้นเป็นส่วนที่ผู้พัฒนาโปรแกรมต้องทำเนื่องจากหน้าจอ นั้นเป็นส่วนสำคัญที่จะเชื่อมต่อกับผู้ใช้งานโปรแกรมหรือก็คือส่วนที่ผู้ใช้งานจะเห็น ลั่งการ และ โต้ตอบได้ ซึ่งจะมีเน้นการออกแบบในทางด้านของรูปแบบหน้าตา บุ่มกด ช่องที่ใช้พิมพ์สำหรับค้นหา ขนาดตัวอักษร สี และรูปภาพ เป็นต้น

การออกแบบหน้าจอสำหรับระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยี การจัดกลุ่มเคปีน (K-Means) จะเน้นในแพลตฟอร์มเว็บแอปพลิเคชันที่เป็นหน้าจอคอมพิวเตอร์หรือ Desktop เป็นหลัก โดยมีรายละเอียดดังนี้

3.5.1 หน้าแรก

The screenshot shows a header with a logo for 'Intern assistant' and a red circle with the number '1'. Below the header is a search bar with a placeholder 'ค้นหาสถานประกอบการ' (Search Business) and a purple 'ค้นหา' (Search) button. The main content area is titled 'ค้นหาสถานประกอบการ' (Search Business) and includes a sub-instruction 'ระบุรายละเอียดความสนใจที่สาขาพิจารณา' (Specify details for branch consideration). A red circle with the number '2' highlights the search bar. Below the search bar is a list of business cards:

- Ernsler - Schmeler**
192 Screeching Street, West Malling, BD72 6RG
Joe@toppid.org
07183 776 242
Online marketing
- Luettgen - Hintz**
400 Kindly Road, Llandover, W27 1TF
Diarney@fuel.co.uk
07223 538 213
Online marketing
- Stehr, Doyle and Schultz**
430 Payment Avenue, Surrey, TS77 7HA
Harley@wealthy.co
07154 387 563
Network
- Lemke and Sons**
467 Chilly Road, Winsford, TA88 2SR
- Champlin - Kessler**
229 Pencil Close, Countess Wear, BN78 8RH
- Christiansen Group**
460 Annoying Crescent, Newton Stewart, PH6 6Q

A red circle with the number '3' highlights the third business card.

ภาพที่ 28 หน้าแรก

จากภาพที่ 28 แสดงการออกแบบหน้าแรกของเว็บไซต์ซึ่งประกอบไปด้วยล้วนประกอบดังนี้
 หมายเลขอ 1 เมนูหลักสามารถเชื่อมโยงไปยังหน้าแสดงรายละเอียดของเว็บไซต์ หน้าสถานประกอบการทั้งหมด และหน้าสถานประกอบการแต่ละประเภท
 หมายเลขอ 2 เป็นช่องสำหรับรับระบุความสนใจของผู้ใช้เพื่อนำไปค้นหากลุ่มบริษัทที่มีความคล้ายคลึงกับความสนใจของผู้ใช้
 หมายเลขอ 3 เป็นการแนะนำบริษัทที่น่าสนใจจากที่อยู่ในฐานข้อมูล

3.5.2 หน้าเกี่ยว กับ

The screenshot shows a header with a logo for 'Intern assistant' and a red circle with the number '1'. Below the header is a section titled 'เกี่ยวกับ' (About) with a detailed description of the system's architecture and data flow.

รูปแบบการดำเนินการ

น่าจะเขียนข้อความว่า 'นี่คือการอธิบายรายละเอียดเกี่ยวกับการทำงานของระบบ'

กระบวนการ ID ของกลุ่มและบันทึกลงฐานข้อมูล

ภาพที่ 29 หน้าเกี่ยว กับ

SERVICE
API Github Document Article

หมายเลขอ 1 คือการอธิบายรายละเอียดเกี่ยวกับระบบและรูปแบบการทำงานในส่วนต่าง ๆ

3.5.3 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

The screenshot shows a search results page for 'Online marketing'. At the top, there is a navigation bar with icons for Intern assistant, search, and account information. Below the navigation bar, the title 'หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด' is displayed. A red circle with the number '1' is overlaid on the left side of the list. The list contains 25 company names, each with a small thumbnail icon and a link to their profile.

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

ภาพที่ 30 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หมายเหตุ 1 แสดงรายชื่อบริษัทที่อยู่ในกลุ่มที่เลือกคัดทั้งหมดและสามารถคลิกเพื่อดูรายละเอียด

3.5.4 หน้าแสดงรายชื่อบริษัททั้งหมด

The screenshot shows a search results page for all companies. At the top, there is a navigation bar with icons for Intern assistant, search, and account information. Below the navigation bar, the title 'สถานประกอบการทั้งหมด' is displayed. A red circle with the number '1' is overlaid on the left side of the list. The list contains 25 company names, each with a small thumbnail icon and a link to their profile.

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

ภาพที่ 31 หน้าแสดงรายชื่อบริษัททั้งหมด

หมายเหตุ 1 รายชื่อบริษัททั้งหมดที่อยู่ในฐานข้อมูลและสามารถคลิกดูข้อมูลบริษัทได้

3.5.5 หน้าแสดงผลลัพธ์รายชื่อบริษัท

ค้นหาสถานประกอบการ
ไม่พบผลลัพธ์ตามเงื่อนไขที่คุณพิมพ์

1 X ค้นหา

2 ! หมวดหมู่: Online marketing

Ernsner - Schmeler	ก่อตั้งมา 10 ปี	Von, Kunde And Stracke	ก่อตั้งมา 10 ปี
Tramp LLC	ก่อตั้งมา 10 ปี	Bode And Sons	ก่อตั้งมา 10 ปี
Gusikowski - Considine	ก่อตั้งมา 10 ปี	Grady, Huels And Runte	ก่อตั้งมา 10 ปี
Dicki, Welch And Rippin	บ้าน	Balistreri Inc	บ้าน
Boyer - Daugherty	สำนักงาน	Cartwright, Goyette And Watsica	สำนักงาน

ภาพที่ 32 หน้าแสดงผลลัพธ์รายชื่อบริษัท

หมายเลขอ 1 ช่องสำหรับระบุความสนใจของผู้ใช้เพื่อค้นหาบริษัทที่คล้ายคลึง

หมายเลขอ 2 ผลลัพธ์รายชื่อบริษัทที่มีความคล้ายคลึงกับความสนใจที่ผู้ใช้ระบุ

3.5.6 หน้าแสดงข้อมูลบริษัท

1 Ernsner - Schmeler
Online marketing

2 467 Chilly Road, Winsford, TA88 2SR
Joe@torpid.org
07183 776 242

3 รายละเอียดของบริษัท
Repellendus laudantium dignissimos deleniti. Officiis et maiores quod veritatis dignissimos voluptatem possimus. Magni tempore sed. Animi eum voluptas dolorum esse amet quisquam. Tempore suscipit animi harum voluptatem impedit. Temporibus sed apertum impedit cum modi. Autem architecto est eveniet cum. Et dolores assumenda numquam qui qui. Debitis et perspicacis ad iste. Minima repudiandae dolor rerum et aut. Atque magni ullam assumenda a consecetur molestias tenetur dolore 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 beatiae quibusdam. Et qui animi qui in voluptas sed.

SERVICE
API Github Document Article

ภาพที่ 33 หน้าแสดงข้อมูลบริษัท

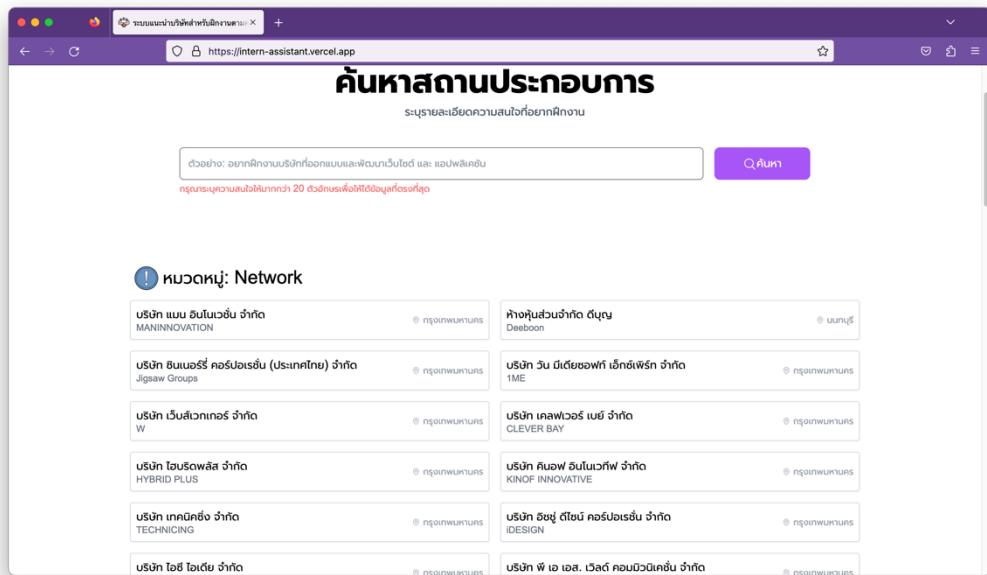
หมายเลขอ 1 แสดงชื่อบริษัท

หมายเลขอ 2 แสดงรายละเอียดข้อมูลที่อยู่ ข้อมูลการติดต่อของบริษัท

หมายเลขอ 3 แสดงข้อมูลรายละเอียดรูปแบบธุรกิจที่บริษัทดำเนินกิจการอยู่

3.6 การใช้งานระบบ

1. ระบุความสนใจในรูปแบบธุรกิจ หรือสิ่งที่อยากรับในการฝึกงานลงในช่องค้นหา
2. เมื่อได้ผลลัพธ์บริษัทที่มีความคล้ายคลึงกับสิ่งที่ค้นหาแล้วผู้ใช้สามารถทำการพิจารณา บริษัทเพื่อเลือกตัดสินใจในการฝึกงานได้



ภาพที่ 34 ตัวอย่างผลลัพธ์จากการค้นหาด้วยความสนใจของผู้ใช้

บทที่ 4

ผลการดำเนินงาน

การทำงานของระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่ม เคเมิน (K-Means) มีขั้นตอนการทำงานดังนี้

4.1 การวิเคราะห์และการตัดคำ (Word segmentation)

4.1.1 การวัดค่าความแม่นยำในการตัดคำ

4.1.2 การจัดกลุ่มข้อมูล

4.1.3 การกำหนดชื่อกลุ่ม

4.1.1 การวัดค่าความแม่นยำในการตัดคำ

ในขั้นตอนการทำ Word segmentation นั้นมีขั้นตอนอยู่ในการทำร่วมด้วยหลายขั้นตอนหลังจากทำความลักษณะของข้อมูล คือการหาคำสำคัญของแต่ละประโยคในที่นี่คือรายละเอียด ธุรกิจของแต่ละบริษัท การกำจัดคำที่ไม่สื่อความหมายหรือ Stop word ออกไปจากประโยคเพื่อให้ได้ ประโยคที่มีเนื้อหาใจความดีที่สุด และอีกขั้นตอนสำคัญคือการหาคีย์เวิร์ดของแต่ละประโยคเพื่อที่จะ ได้ทราบว่าประโยคนั้น ๆ กำลังสื่อถึงเรื่องไหนเป็นสำคัญด้วยเทคนิค TF-IDF ที่เป็นการหาหนังสือ ของคำนั้น ๆ ในประโยค

ในการทำงานประมวลผลเกี่ยวกับการประมวลผลภาษาธรรมชาติจำเป็นต้องมีการ ตัดคำออกเป็นคำ ๆ เพื่อจะได้ง่ายและนำไปเข้าสู่กระบวนการต่าง ๆ ได้อย่างง่ายโดยปกติแล้วการตัด คำในภาษาอังกฤษนั้นสามารถตัดได้โดยใช้การเว้นวรรคเป็นเงื่อนไขในการตัด แต่ในภาษาไทยนั้นการ เขียนนั้นไม่ได้มีการเว้นวรรคคำให้มีอนุญาติอย่างกตัญหากทำการตัดคำนั้นจะใช้เว้นวรรคมาตัดคำลดอด ไม่ได้ จำเป็นต้องใช้อัลกอริทึมอื่น ๆ เช่นมาช่วย เช่นการใช้ Dictionary-based, Maximum Matching เป็นต้น

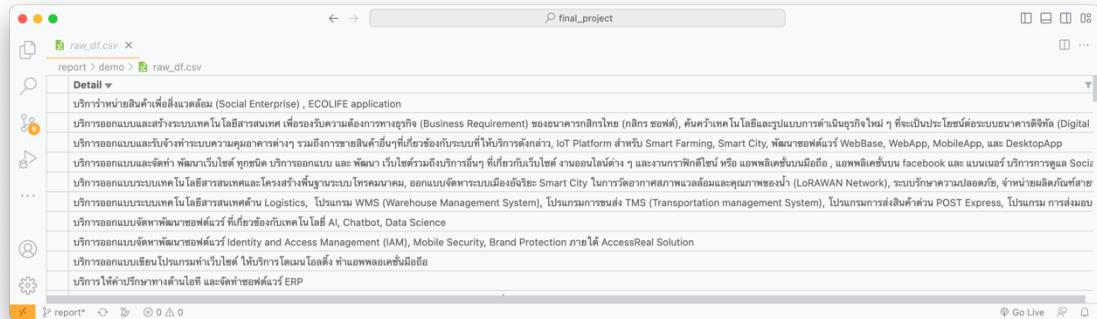
และในภาษา Python มีเครื่องมือที่สามารถช่วยอำนวยความสะดวกในการตัดคำ ภาษาไทยอย่าง Pythainlp ที่ผู้วิจัยได้เลือกใช้ในโครงการนี้ในไลบรารีนั้นสามารถตัดคำได้หลาย Engine ด้วยกันและในแต่ละตัวเลือกซึ่งใช้อัลกอริทึมต่างกันยกตัวอย่างเช่น newmm, longest, newmm-safe, mm, icu, deepcut, attacut เป็นต้น

1. การวัดค่าความแม่นยำในการตัดคำ

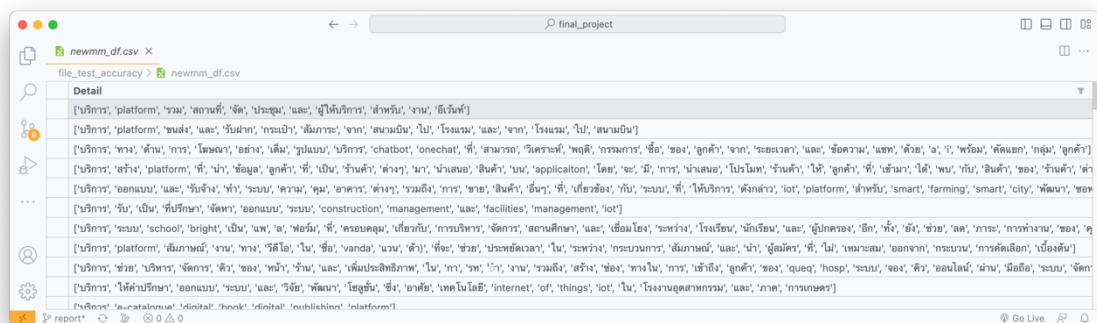
การเลือก Engine มาใช้ตัดคำจำเป็นต้องมีการวัดค่าความถูกต้องเพื่อที่จะได้ผลลัพธ์ที่ ตรงกับความต้องการมากที่สุดและในโครงการนี้ผู้วิจัยได้เลือก Engine มาทดสอบด้วยกันจำนวน 3 ตัวเลือกดังนี้

1. newmm – dictionary-based, Maximum Matching + Thai Character Cluster
2. deepcut – wrapper for DeepCut, learning-based approach
3. longest – dictionary-based, Longest Matching

วิธีทดสอบคือทำการสุ่มเลือกประโภคมาจำนวน 100 ประโภค ทำการตัดคำในแต่ละประโภคเองโดยไม่ใช้ตัวช่วย จากนั้นใช้ Engine ในไลบรารี Pythainlp เพื่อตัดคำและนำมาเทียบกับประโภคที่ผู้วิจัยตัดให้โดย Engine ให้เห็นว่าความเหมือนกับที่ผู้วิจัยตัดไว้มากที่สุดจะถือว่ามีความใกล้เคียงกับความต้องการของผู้วิจัยมากที่สุด



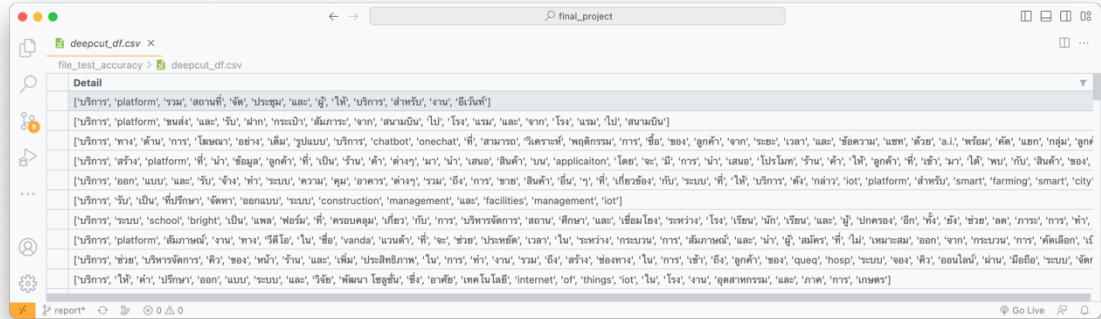
ภาพที่ 35 ตัวอย่างข้อมูลต้นฉบับ



ภาพที่ 36 ตัวอย่างการตัดคำโดยใช้ Engine newmm



ภาพที่ 37 ตัวอย่างการตัดคำโดยใช้ Engine longest



ภาพที่ 38 ตัวอย่างการตัดคำโดยใช้ Engine deepcut

ยกตัวอย่างการทำการทำหัดล่องการตัดคำโดยใช้ Engine newmm, longest และ deepcut โดยที่ใช้วิธีเทียบการทำหัดคำโดยที่ผู้วิจัยทำการตัดคำเองและนำไปเทียบกับการทำให้แล้วโดยนับจำนวนเฉพาะคำที่ให้แล้วกับหัดคำโดยเด่นชัดนี้จึงถือว่า Engine ได้มีการทำหัดคำที่ตรงกับผู้วิจัยมากที่สุด

โจทย์: บริการ Software Business Solutions ระบบรับสมัครบุคลากรออนไลน์ ระบบลงทะเบียนเรียนออนไลน์

ตาราง 15 ตารางตัวอย่างการวัดค่าความแม่นยำในการตัดคำ

ลำดับ	Human	Newmm	Longest	Deepcut
1	บริการ	บริการ	บริการ	บริการ
2	software	software	software	software
3	Business	business	business	business
4	Solutions	solutions	solutions	solutions
5	ระบบ	ระบบ	ระบบ	ระบบ
6	รับสมัคร	รับสมัคร	รับสมัคร	รับ
7	บุคลากร	บุคลากร	บุคลากร	สมัคร
8	ออนไลน์	ออนไลน์	ออนไลน์	บุคลากร
9	ระบบ	ระบบ	ไลน์	ออนไลน์
10	ลงทะเบียนเรียน	ลงทะเบียนเรียน	ระบบ	ระบบ
11	ออนไลน์	ออนไลน์	ลงทะเบียนเรียน	ลง
12			ออนไลน์	ลงทะเบียน
13				เรียน
14				ออนไลน์
รวม		11	10	9

จากตารางที่ 15 จะเห็นว่าเมื่อทำการใช้ไลบรารีในการตัดคำแล้วสามารถสรุปได้ว่า Engine ที่มีความคล้ายกับผู้วิจัยมากที่สุดคือ newmm รองลงมาคือ longest และ deepcut

ตาราง 16 ผลการทดสอบความแม่นยำการตัดคำ

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

จากการทดลองของจากข้อมูลประมาณ 100 รายการพบว่า Engine newmm ในไลบรารี Pythainlp นั้นมีความแม่นยำกับที่ผู้วิจัยตัดมากที่สุดที่ 90.99% รองลงมาที่ longest ความแม่นยำอยู่ที่ 83.04% และความแม่นยำน้อยที่สุดคือ deepcut ที่มีความแม่นยำ 76.65% ดังภาพที่ 39 ดังนั้นโครงการนี้จึงใช้ Engine newmm ในการตัดคำเพื่อนำไปประมวลผลต่อในขั้นตอนอื่น ๆ

```
onze@Tinngrits-MacBook-Pro:~/Documents/final_project
..final_project (-zsh)  #1 ..final_project (-zsh)  #2 +
python calculate_accuracy.py
newmm: 90.99%
deepcut: 76.65%
longest: 83.04%
```

ภาพที่ 39 ผลการวัดความแม่นยำในการตัดคำของ Engine ในไลบรารี Pythainlp

4.1.2 การคำนวณค่า TF-IDF

เมื่อได้เครื่องมือที่จะช่วยตัดคำแล้วขั้นตอนต่อไปคือการนำมาทำการจัดกลุ่มข้อมูลโดยเทคนิคที่เลือกใช้คือการหาค่า TF-IDF (Term Frequency–Inverse Document Frequency) เพื่อหาว่าคำไหนในประโยคนั้นเป็นคำสำคัญของประโยคนั้น ๆ โดยการวัดจากน้ำหนักของคำด้วยวิธีดังกล่าว

	aaa	ab	abap	abeam	ablerex	abroad	academic	acceptance	access	accessories	...	ໄອທີ	ໄອສ ຕຽນ	ໄອ ເຕັມ	ໄອ ແພດ	ໄອ ໂນບາຍ	ໄອ ໄກ	ໄອ ໂຕ	ໄອໂຄຣເຊນ
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	
1645 rows x 8526 columns																			

ภาพที่ 40 ตัวอย่างตาราง TF-IDF แสดงน้ำหนักของคำ

เมื่อได้ชุดของประโยชน์ที่ทำการตัดคำเรียบร้อยแล้วจึงนำเข้าสู่กระบวนการลบคำที่เป็น Stop word หรือคำที่ไม่มีความหมายอوكและตัวอักษรพิเศษต่าง ๆ ด้วยฟังก์ชันในไลบรารี Pythainlp และ nltk ยกตัวอย่างเช่น

ประโยชน์ต้น: บริการ Platform รวมสถานที่จัดประชุมและผู้ให้บริการสำหรับงานอีเว้นท์ที่ติดต่อสัมภาษณ์ไม่ใช่ความหมาย (Stop word): รวมจัดและที่ที่สุด

ประโยชน์ใหม่: บริการPlatform/สถานที่ประชุมผู้ให้บริการสำหรับงานอีเว้นท์ได้

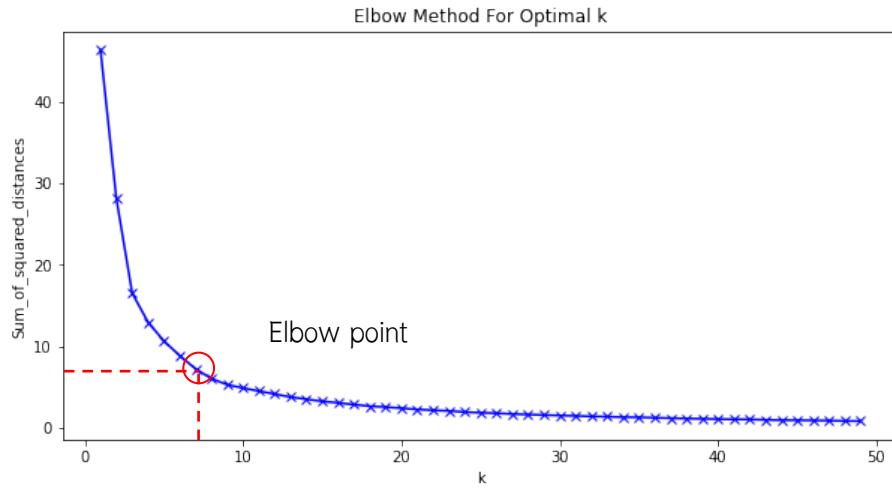
```

onze@Tinngrits-MacBook-Pro:~/Documents/final_project
..final_project (~zsh) ❶ t1 ..final_project (~zsh)
└─ python report/demo/main.py
Default: ນິກາຮ່າ Platform ລວມສານທີ່ຈັດປະຊຸມແລະຜູ້ໃຫ້ບໍລິການສໍາຫວັນນີ້ເວັນທີ່ສຸດ
Keep stop word: ['ນິກາຮ່າ', 'Platform', 'ຮ່ວມ', 'ສອນທີ່', 'ຈົດ', 'ປະຊຸມ', 'ແລະ', 'ຜູ້ໃຫ້ບໍລິການ', 'ສໍາຫວັນ', 'ຈານ', 'ເອົ້ວນທີ່', 'ທີ່', 'ດີ', 'ທີ່ສຸດ']
Removed stop word: ['ນິກາຮ່າ', 'platform', 'ສອນທີ່', 'ປະຊຸມ', 'ຜູ້ໃຫ້ບໍລິການ', 'ສໍາຫວັນ', 'ຈານ', 'ເອົ້ວນທີ່', 'ດີ']
~/Documents/final_project report -t1 !2 ?2

```

ภาพที่ 41 ตัวอย่างการตัดคำและลบ Stop word

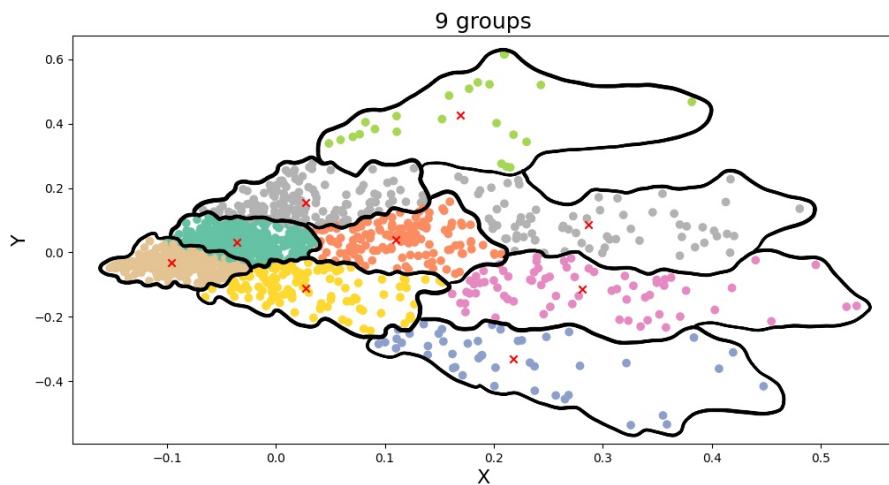
การจัดกลุ่มของข้อมูลหรือการทำ Text clustering นั้นโครงงานนี้จะใช้เทคนิค K-Means มาใช้ในการจัดกลุ่มโดยเลือกจำนวนกลุ่มจากการทำ Elbow method เพื่อหาจำนวนกลุ่มที่ดีที่สุดโดยใช้ Elbow method ในการหาจำนวนกลุ่มที่เหมาะสมจะอยู่ที่บริเวณส่วนโคงคล้ายข้อคอกในที่นี่จะประมาณกลุ่มได้ 6-9 กลุ่ม



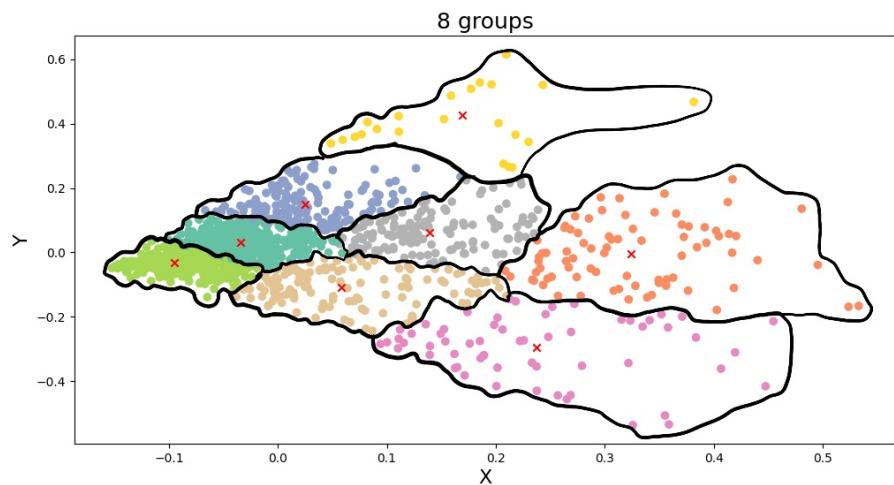
ภาพที่ 42 การทำ Elbow method

จากนั้นทำการทดลองจัดกลุ่มข้อมูลด้วยจำนวนกลุ่มที่แตกต่างกันแต่ข้อมูลเดียวกัน ทำการทดลองจัดกลุ่มด้วย 4 กรณีได้ดังนี้

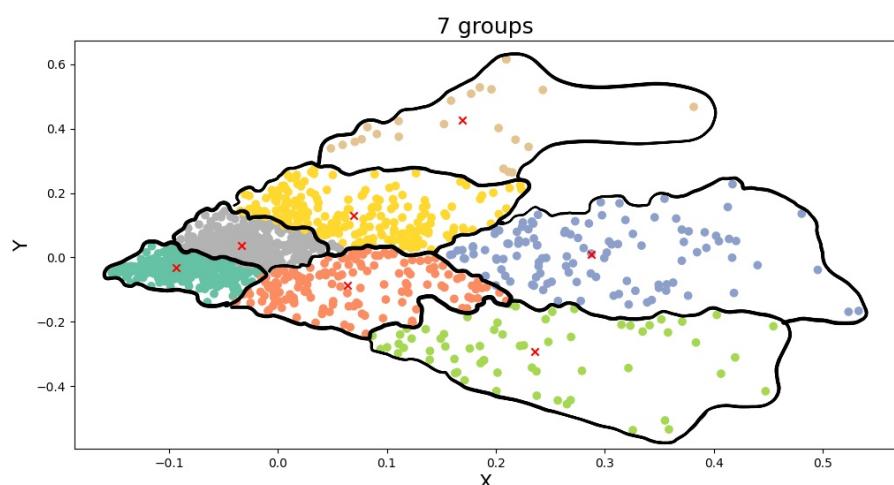
- จำนวน 9 กลุ่ม พบร้าข้อมูลของข้อมูลค่อนข้างแคบมากและมีกลุ่มที่มีเนื้อหาซ้ำกันมากกว่า 1 กลุ่ม
- จำนวน 8 กลุ่ม พบร้าข้อมูลค่อนข้างแคบและเนื้อหาบางล้วนจะปนอยู่ในกลุ่มอื่น ๆ ที่มีใกล้กัน
- จำนวน 7 กลุ่ม พบร้าข้อมูลค่อนข้างดีและเนื้อหาในกลุ่มนั้นมีปะปนกันน้อยมากและไม่มีกลุ่มที่มีเนื้อหาซ้ำกัน
- จำนวน 6 กลุ่ม พบร้าข้อมูลของข้อมูลกว้างมากและทำให้ใน 1 กลุ่มนั้นมีเนื้อหาที่มากกว่า 1 อย่างทำให้ไม่สามารถระบุแน่ชัดได้ว่าเกี่ยวกับเรื่องใดเป็นหลัก



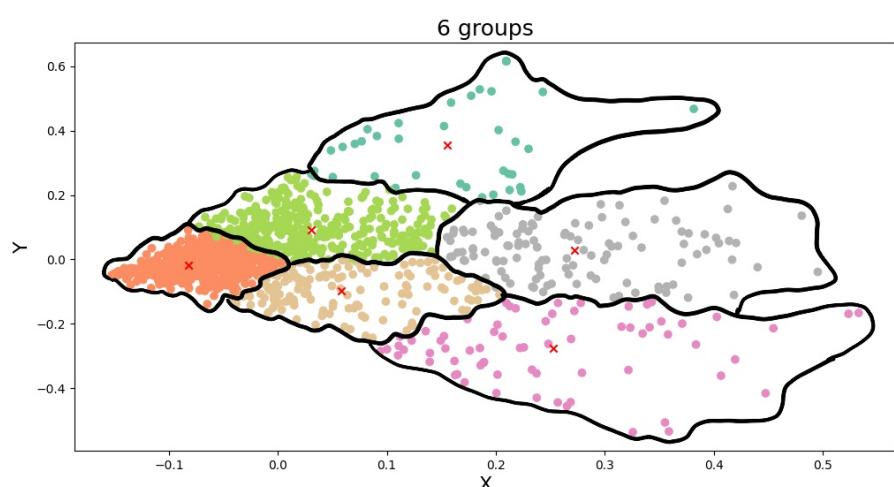
ภาพที่ 43 จัดกลุ่มข้อมูลจำนวน 9 กลุ่ม



ภาพที่ 44 จัดกลุ่มข้อมูลจำนวน 8 กลุ่ม



ภาพที่ 45 จัดกลุ่มข้อมูลจำนวน 7 กลุ่ม



ภาพที่ 46 จัดกลุ่มข้อมูลจำนวน 6 กลุ่ม

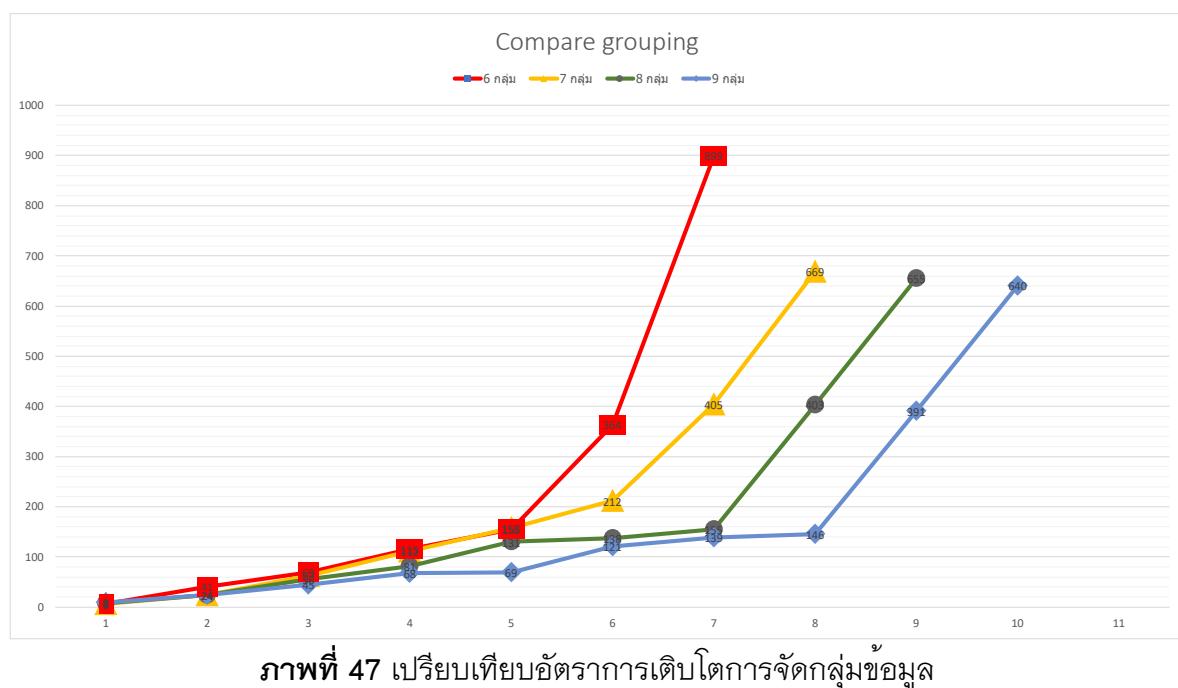
ทำการทดลองนับจำนวนรายการบริษัทแต่ละประเภทในแต่ละกลุ่มเพื่อศูนย์เฉลี่ยและการ
กระจายตัวของข้อมูลบริษัท

ตาราง 17 แสดงการนับจำนวนบริษัทแต่ละประเภทในการจัดกลุ่มทั้งหมด 1,643 รายการ

จำนวนกลุ่ม	ประเภทของบริษัท	ค่าเฉลี่ย
6	<ul style="list-style-type: none"> - กลุ่ม 1 มี 115 รายการ - กลุ่ม 2 มี 899 รายการ - กลุ่ม 3 มี 69 รายการ - กลุ่ม 4 มี 155 รายการ - กลุ่ม 5 มี 41 รายการ - กลุ่ม 6 มี 364 รายการ 	469.43 รายการ
7	<ul style="list-style-type: none"> - กลุ่ม 1 มี 405 รายการ - กลุ่ม 2 มี 24 รายการ - กลุ่ม 3 มี 112 รายการ - กลุ่ม 4 มี 158 รายการ - กลุ่ม 5 มี 669 รายการ - กลุ่ม 6 มี 63 รายการ - กลุ่ม 7 มี 212 รายการ 	234.71 รายการ
8	<ul style="list-style-type: none"> - กลุ่ม 1 มี 403 รายการ - กลุ่ม 2 มี 81 รายการ - กลุ่ม 3 มี 655 รายการ - กลุ่ม 4 มี 155 รายการ - กลุ่ม 5 มี 138 รายการ - กลุ่ม 6 มี 24 รายการ - กลุ่ม 7 มี 131 รายการ - กลุ่ม 8 มี 56 รายการ 	205.38 รายการ
9	<ul style="list-style-type: none"> - กลุ่ม 1 มี 391 รายการ - กลุ่ม 2 มี 69 รายการ - กลุ่ม 3 มี 640 รายการ - กลุ่ม 4 มี 68 รายการ - กลุ่ม 5 มี 45 รายการ - กลุ่ม 6 มี 121 รายการ - กลุ่ม 7 มี 139 รายการ - กลุ่ม 8 มี 146 รายการ 	182.56 รายการ

ตาราง 17 (ต่อ)

จำนวนกลุ่ม	ประเภทของบริษัท	ค่าเฉลี่ย
9	- กลุ่ม 1 มี 115 รายการ	469.43 รายการ



ภาพที่ 47 เปรียบเทียบอัตราการเติบโตการจัดกลุ่มข้อมูล

จากภาพที่ 43–46 จะเห็นได้ว่าการกระจายตัวของข้อมูลนั้นมีความแตกต่างกันสังเกตได้จากสีที่ระบุตำแหน่งของกลุ่มในแต่ละภาพโดยเฉพาะเมื่อเปรียบเทียบการแบ่งกลุ่มที่ 6 และ 8 จะเห็นว่าขอบเขตของข้อมูลของการแบ่ง 6 กลุ่มนั้นมีความกว้างมากและห่างไกลจากจุดกึ่งกลางของข้อมูล (Centroids point) จึงประมาณได้ว่าการแบ่งกลุ่มที่จำนวน 6 กลุ่มนั้นอาจไม่ได้ประสิทธิภาพความแม่นยำมากพอก และในตรงกันข้ามการแบ่งกลุ่มข้อมูลที่ 9 กลุ่มนั้นจะเห็นได้ว่าขอบเขตของข้อมูลนั้นแคบมากจนถึงทับซ้อนกันในแต่ละกลุ่มถึงแม่ขอบเขตของข้อมูลจะอยู่ใกล้จุดกึ่งกลางของข้อมูล ดังนั้นผู้วิจัยจึงตัดตัวเลือกการแบ่งกลุ่มข้อมูลที่ 6 กลุ่มออกเหลือเพียง 7 8 และ 9 กลุ่ม

และจากภาพ 47 จะเห็นได้ว่าการเติบโตของข้อมูลเมื่อนำมาปริมาณบริษัทในแต่ละประเภทมาเรียงจากน้อยไปมากนั้น อัตราการเติบโตของการจัดกลุ่มข้อมูลที่ 7 กลุ่มนั้นมีอัตราการเติบโตที่คงที่มากที่สุด

จากนั้นทดลองทำการสุมเรียกข้อมูลในแต่ละกรณีของมาเพื่อประกอบการตัดสินใจที่จะเลือกจำนวนกลุ่มของข้อมูล

จากการทดลองสุมเรียกข้อมูลหลาย ๆ ครั้งพบว่าจำนวนของการจัดกลุ่มที่มีค่าข้อมูลทับซ้อนกันน้อยที่สุดอยู่ที่ 7 กลุ่มทำให้ผู้วิจัยเลือกที่จะแบ่งกลุ่มข้อมูลที่ 7 กลุ่มแต่ทั้งนี้ก็ยังมีข้อมูลที่ทับซ้อนกันอยู่บ้างเล็กน้อยซึ่งอยู่ในระดับที่รับได้ และเมื่อทำการแบ่งกลุ่มข้อมูลเรียบร้อยแล้วจึงสามารถบันทึกข้อมูลพร้อมกับ ID ของกลุ่มเพื่อนำไปใช้ฐานข้อมูลและใช้งานต่อไป

4.1.3 การกำหนดชื่อกลุ่ม

เมื่อได้ข้อมูลที่สมบูรณ์อยู่ในฐานข้อมูลแล้วนั้นการแสดงผลข้อมูลของกลุ่มจากหน้าเว็บไซต์จำเป็นต้องมีการตั้งชื่อกลุ่มนี้องจากข้อมูลที่ได้จากการจัดกลุ่มคือ ID ซึ่งคือตัวเลขตั้งแต่ 0-6 เนื่องจากทำการกำหนดจำนวนกลุ่มไว้ที่ 7 กลุ่มดังนั้นเพื่อให้การแสดงผลในหน้าเว็บไซต์และให้การเรียกกลุ่มง่ายขึ้น จึงทำการตั้งชื่อกลุ่มโดยชื่อจะอยู่ในประเภทของเทคโนโลยีเนื่องจากข้อมูลที่มีนั้นเป็นบริษัทที่ทำเกี่ยวกับอุตสาหกรรมเทคโนโลยีหรือบริษัททางด้านไอที การตั้งชื่อของกลุ่มข้อมูลนั้นได้ทำการอ้างอิงซึ่งจากประเทงานโอดีจากเว็บไซต์ th.jobsdb.com เป็นหลัก ซึ่งมีอยู่ 18 ประเภท รายชื่อประเทงานโอดีที่มีในเว็บไซต์ th.jobsdb.com มีดังนี้

1. งาน Application Network
2. งาน Software
3. งาน Database
4. นักวิทยาศาสตร์ข้อมูล
5. งาน Hardware
6. งาน IT Audit
7. งานปรึกษาโอดี
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. งานโอดีอื่น ๆ

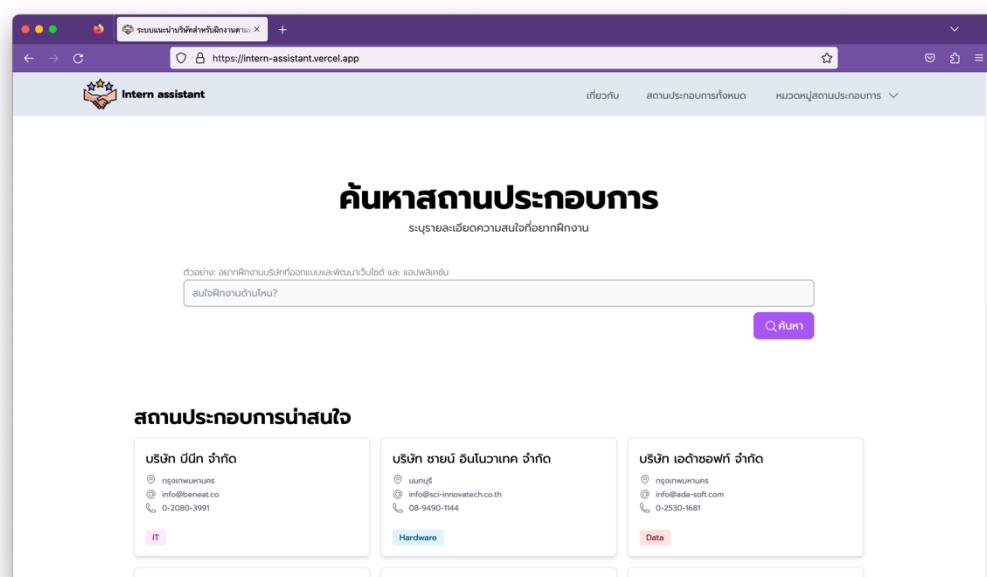
แต่ในโครงงานนี้มีกลุ่มข้อมูลเพียง 7 กลุ่มดังนั้นจึงต้องเลือกประเทงานที่ตรงกับข้อมูลในกลุ่มมากที่สุดเท่านั้น โดยวิธีที่ใช้เลือกคือการสุ่มข้อมูลในแต่ละกลุ่มตั้งแต่ 0-6 และตรวจดูว่าควรจะได้ชื่อกลุ่มเป็นประเทงานไหน

จากการทำการสุ่มเรียกข้อมูลดูทั้ง 7 กลุ่ม รายชื่อประเทงานที่สามารถใช้ตั้งชื่อกลุ่มข้อมูลโดยอ้างอิงของเว็บไซต์ th.jobsdb.com มีดังนี้

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

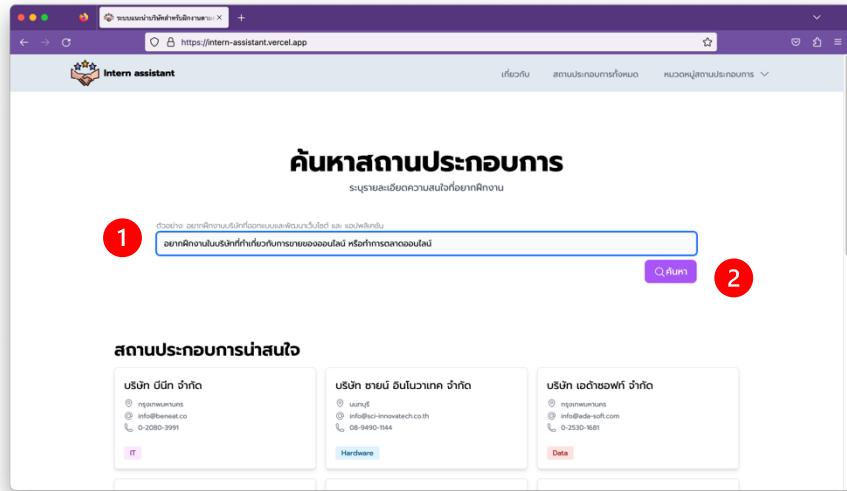
4.2 ขั้นตอนการใช้งานสำหรับผู้ใช้งาน

4.2.1 หน้าแรกเว็บไซต์ Intern-assistant



ภาพที่ 48 หน้าแรกเว็บไซต์ Intern-assistant

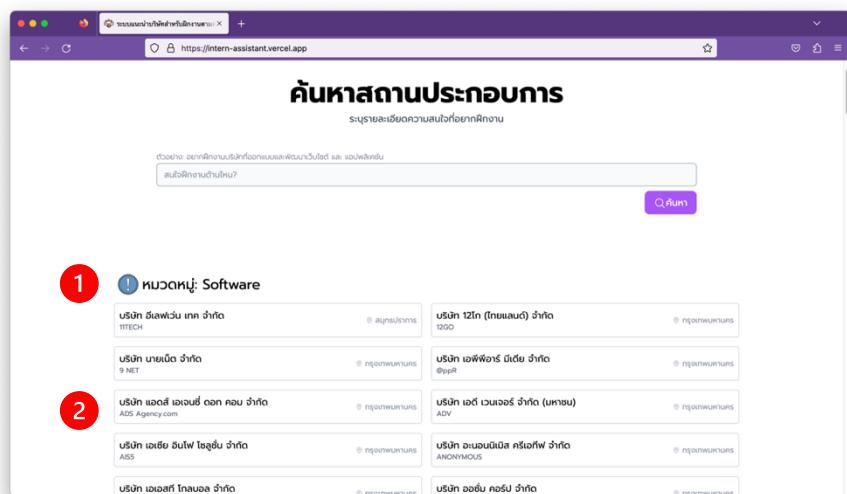
4.2.2 គន្លាបរិម័ត្ត



រាងទី 49 គន្លាបរិម័ត្ត

ឃមមិលី 1 ពិមិលីគន្លាបរិម័ត្តទីតែងការដឹកការណ៍តាមរាយលក់អីយុទ្ធម៌សន្និច្ចូបប្រួចកិច
ឃមមិលី 2 កតុប្រុមគន្លាប

4.2.3 អាជីវកម្មផលិតផលការគន្លាប

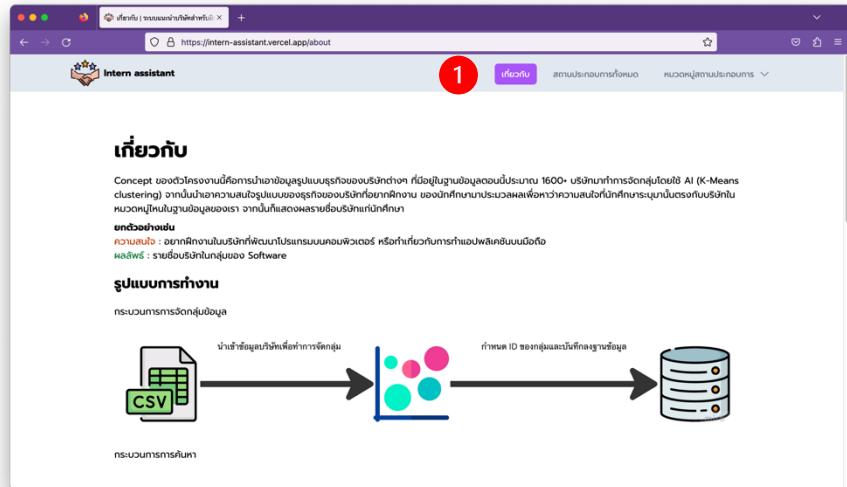


រាងទី 50 អាជីវកម្មផលិតផលការគន្លាប

ឃមមិលី 1 ឱ្យឈឺក្នុងក្នុងក្នុង

ឃមមិលី 2 រាយឱ្យឈឺបរិម័ត្តទាំងអស់ក្នុងក្នុង

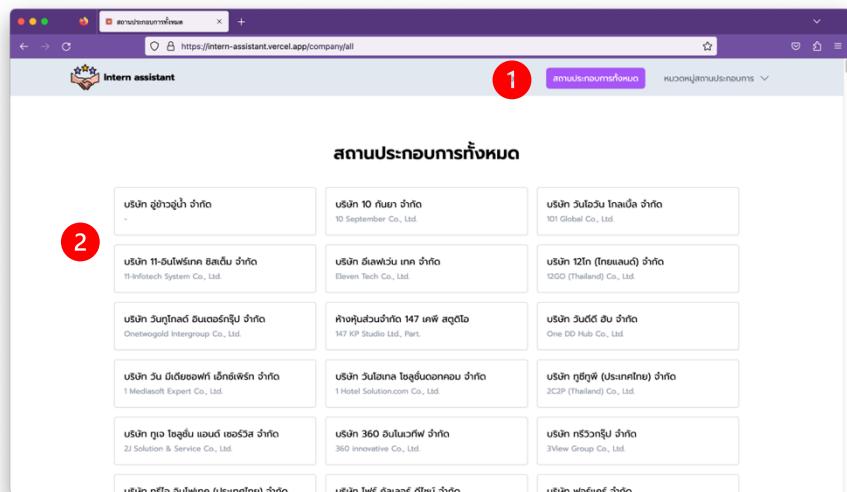
4.2.4 หน้าเกี่ยวกับ



ภาพที่ 51 หน้าเกี่ยวกับ

หมายเลขอ 1 เมนูหน้าเกี่ยวกับแสดงข้อมูลเกี่ยวกับระบบและการทำงาน

4.2.5 หน้าแสดงรายชื่อบริษัททั้งหมด



ภาพที่ 52 หน้าแสดงรายชื่อบริษัททั้งหมด

หมายเลขอ 1 เมนูหน้ารายชื่อบริษัททั้งหมด

หมายเลขอ 2 รายชื่อบริษัททั้งหมด

4.2.6 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

ภาพที่ 53 หน้าแสดงรายชื่อบริษัทในกลุ่มทั้งหมด

หมายเลขอ 1 เมนูรายชื่อกลุ่มบริษัท

หมายเลขอ 2 รายชื่อกลุ่มบริษัท

หมายเลขอ 3 รายชื่อบริษัทในกลุ่ม

4.2.7 หน้ารายละเอียดบริษัท

1 บริษัท ดิฟเฟอร์ชีต จำกัด
DifferSheet Co., Ltd. (DifferSheet)
Data

2 บริษัท ดิฟเฟอร์ชีต จำกัด
89/13 หมู่ที่ 5 ถนนนา-{ชื่อ} บ้านนา-{ชื่อ} ตำบล-{ชื่อ} อำเภอ-{ชื่อ} จังหวัด-{ชื่อ}
อุบลราชธานี 10540
hello@differsheet.com
0-2130-6525
http://www.differsheet.com

3 รายละเอียดบริษัท
บริการและแพลตฟอร์มสำหรับเขียนเรียนบันทึกในส่วนตัว (เช่น Gamification หลักสูตร Social Network และหลักสูตรภาษาไทย) หรือ
เขียนเรียนบันทึกในส่วนตัว (เช่น Gamification หลักสูตร Social Network และหลักสูตรภาษาไทย) หรือ

ภาพที่ 54 หน้ารายละเอียดบริษัท

4.3 การวัดค่าความคล้ายคลึง

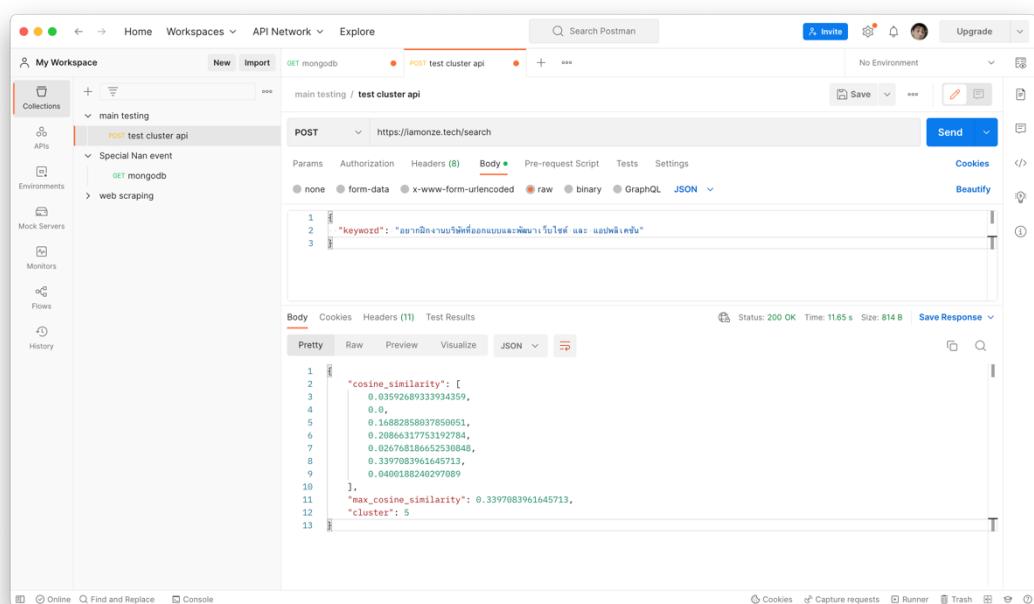
Cosine similarity เป็นเทคนิคที่นำมาใช้หาความคล้ายคลึงระหว่างความสนใจของผู้ใช้และข้อมูลบริษัทที่อยู่ในฐานข้อมูลยิ่งค่า Cosine similarity เข้าใกล้ 1 表示ว่าประโยชน์มีความคล้ายคลึงกับข้อมูลบริษัทในกลุ่มนั้นมากดังภาพที่ 55

```
onze@Tinngrits-MacBook-Pro:~/desktop/final_project
..final_project (-zsh) ⑥ 1 ..final_project (-zsh) ⑥ 2 + ~/Desktop/final_project
python cosine_similarity.py
Keyword: ออกรอบเว็บไซต์ด้วย react js ทำเกี่ยวกับการเขียนเว็บ การตลาดออนไลน์ด้วย และ SEO
cluster: 6
cosine similarity: 0.07871941760441518
~/desktop/final_project develop !3 ?2 7s 12:56:32
```

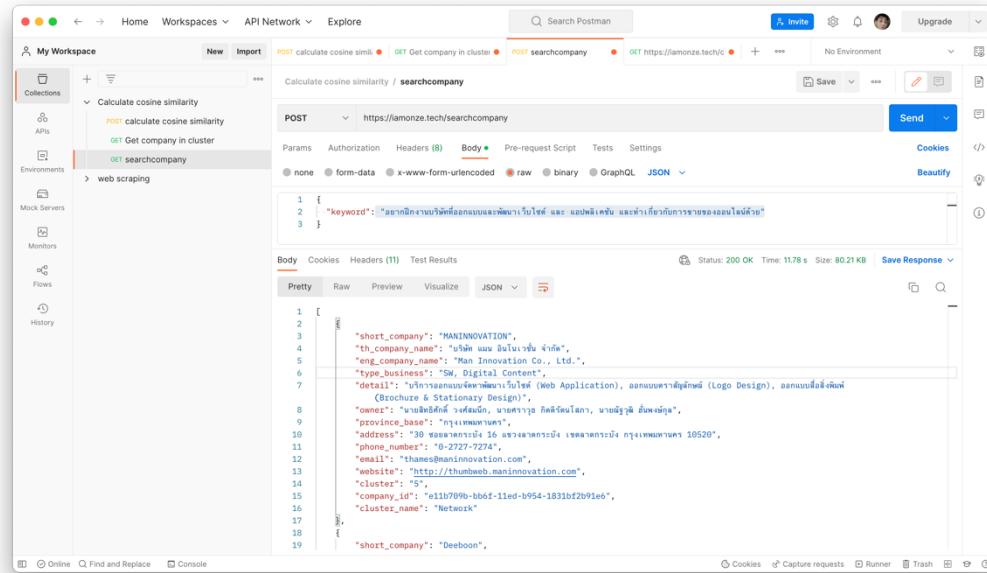
ภาพที่ 55 ตัวอย่างการคำนวณค่า Cosine similarity

4.3.1 การหากลุ่มที่มีความคล้ายคลึงมากที่สุด

- คำนวณค่า Cosine similarity ทุกกลุ่มจากความสนใจของผู้ใช้ผ่าน API ที่สร้างไว้เพื่อคำนวณโดยเฉพาะ



ภาพที่ 56 ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าความคล้ายคลึง



ภาพที่ 57 ตัวอย่างการคำนวณค่า Cosine similarity ผ่าน API และคืนค่าเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่คล้ายที่สุด

บทที่ 5

สรุปผลการดำเนินงาน

การพัฒนาระบบแนะนำสถานที่พิจารณาด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเครื่อง (K-Means) ในครั้งนี้สามารถสรุปการดำเนินงาน ปัญหาที่เกิดขึ้นระหว่างการดำเนินงาน ข้อเสนอและแนวทางพัฒนาต่อไปดังนี้

5.1 สรุปผลการดำเนินงาน

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

การจัดกลุ่มข้อมูลทั้งหมดจำนวน 4 กลุ่มได้แก่การจัดกลุ่มข้อมูลที่ 6 7 8 และ 9 กลุ่มได้ผลสรุปดังนี้ กรณีแบ่งกลุ่มข้อมูลที่ 6 กลุ่มนั้นมีอสูรเมืองเรียกข้อมูลดูแล้วพบว่าขอบเขตข้อมูลนั้นกว้าง เกินไปและมีข้อมูลทับซ้อนกันจำนวนมาก กรณีแบ่งข้อมูลที่ 7 กลุ่มพบว่ามีอสูรเมืองเรียกดูข้อมูล ข้อมูล มีความทับซ้อนกันน้อยมากและขอบเขตของข้อมูลก็อยู่ในระดับที่เหมาะสมอยอมรับได้ กรณีแบ่งกลุ่มที่ 8 กลุ่มพบว่ามีความคล้ายคลึงกับการแบ่งกลุ่มที่ 7 กลุ่มแต่ขอบเขตของข้อมูลบางกลุ่มนั้นแคบ เกินไปทำให้มีเนื้อหาที่ซ้ำกันกับกลุ่มอื่น และกรณีแบ่งกลุ่มที่ 9 กลุ่มพบว่าขอบเขตของข้อมูลนั้นแคบ ที่สุดและแต่ละกลุ่มนั้นมีความทับซ้อนกันค่อนข้างมากจึงเกิดกลุ่มที่มีเนื้อหาแบบเดียวกันแต่อยู่คนละกลุ่ม

ซึ่งจะเห็นได้ว่ากรณีการแบ่งกลุ่มที่ 6 และ 9 กลุ่มนั้นขอบเขตของเนื้อหานั้นอยู่ในระดับที่ไม่ค่อยติดกันเมื่อเทียบกับการแบ่งกลุ่มที่ 7 และ 8 กลุ่ม และในการทดลองวิธี Elbow method จุดที่อยู่ตรงมุมมองระหว่างจุดที่ 7-8 ผู้วิจัยจึงเลือกเปรียบเทียบกันและหาข้อสรุปได้ว่าเลือกแบ่งกลุ่มที่ 7 กลุ่มเป็นการแบ่งกลุ่มข้อมูลที่มีประสิทธิภาพมากที่สุดจากการสุ่มเรียกดูข้อมูลของแต่ละกรณีเพื่อมาใช้ในการพัฒนาระบบ

ในการหาค่าความคล้ายคลึงของข้อมูลด้วยวิธี Cosine similarity นั้นผลลัพธ์การคำนวณและการวิเคราะห์เพื่อหากลุ่มที่เหมาะสมกับความสนใจของผู้ใช้นั้นพบว่าเมื่อได้กลุ่มจากที่การคำนวณแล้วนั้นรูปแบบธุรกิจค่อนข้างตรงกับความสนใจที่ผู้ใช้สั่งมา

5.1.1 จุดเด่นของระบบ

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

5.2 สรุปปัญหาที่เกิดระหว่างการดำเนินงาน

5.2.1 ในการที่จะเพิ่มข้อมูลบริษัทลงในฐานข้อมูลเพิ่มจำเป็นต้องทำการ Word segmentation ข้อมูลใหม่ที่จะเข้ามาร่วมกับข้อมูลเดิมที่มีอยู่แล้วดังนั้นข้อมูลของแต่ละบริษัทก็จะถูกเปลี่ยนกลุ่มไปทุกครั้งทั้งมีการเพิ่มข้อมูลใหม่

5.2.2 เมื่อทำการ Clustering ข้อมูลใหม่แล้วต้องทำการตั้งชื่อให้กับกลุ่มข้อมูลใหม่ เพราะเมื่อมีข้อมูลที่เปลี่ยนไปเนื้อหาบริษัทในกลุ่มเดิมก็อาจเปลี่ยนไปยกตัวอย่างเช่น กลุ่มที่ 0 เดิมเป็นกลุ่มของ Network แต่เมื่อมีการเพิ่มข้อมูลใหม่และทำการ Clustering ใหม่กลุ่ม 0 ก็อาจจะกลายเป็น Data เพราะเนื้อหาในกลุ่มนั้นเปลี่ยนไป หรืออาจมีกลุ่มเพิ่มเติมขึ้นมากนอกเหนือจากปัจจุบัน

5.2.3 ข้อมูลที่ได้รับมาเมื่อมีคำที่สะกดผิดจนไม่สามารถแก้ไขได้ด้วยผู้พัฒนาเอง การทำ Clustering นั้นอาจจะไม่ได้แยกแยะข้อมูลได้ดียกตัวอย่างเช่น ประโยชน์ที่มีคำว่า “แอปพลิเคชัน” ที่สะกดถูกต้องอาจจะอยู่คู่คละกลุ่มกับประโยชน์ที่มีคำว่า “แอพพลิเคชัน” ที่มีการสะกดผิด

5.3 แนวทางพัฒนาระบบในอนาคต

5.3.1 พัฒนาความแม่นยำในการจัดกลุ่มข้อมูล

5.3.2 พัฒนาให้สามารถแนะนำตำแหน่งงานในบริษัทได้

5.3.3 เพิ่มชุดข้อมูลให้มากขึ้นเพื่อเพิ่มความแม่นยำของเว็บไซต์

5.3.4 พัฒนาความเร็วของอัลゴริทึมในการคำนวณความคล้ายคลึง (Cosine similarity)

5.3.5 พัฒนาขั้นตอนการเรียกดูข้อมูลให้ง่ายขึ้น

5.3.6 พัฒนาให้รองรับการกรองตัวเลือกที่จะค้นหา

5.3.7 พัฒนาเว็บไซต์ให้มีความปลอดภัยมากขึ้น

5.4 แบบประเมินความพึงพอใจของผู้ใช้

จากการประเมินความพึงพอใจของผู้ใช้ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) (Computer Internship Recommendation System With K-Means Clustering) สามารถสรุปผลได้ดังนี้

การกำหนดเกณฑ์การพิจารณา

เกณฑ์การพิจารณาระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) 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/ความเข้าใจตีมาก
 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 18 แสดงจำนวนค่าเฉลี่ยความพึงพอใจต่อระบบ

รายละเอียด	ระดับความพึงพอใจ/ความเข้าใจ						เกณฑ์การประเมิน
	5	4	3	2	1	ค่าเฉลี่ย	
มีการออกแบบหน้าจอสำหรับผู้ใช้งานอย่างเหมาะสม	7	6	3	1	0	4.12	ดี
วิธีการใช้งานง่ายต่อการทำความเข้าใจ	6	7	3	1	0	4.06	ดี
ประสิทธิภาพความเสถียรในการทำงานของเว็บแอปพลิเคชัน	2	9	6	0	0	3.76	ดี
ความเหมาะสมของรูปแบบของหน้าจอ แจ้งเตือนต่างๆ	5	7	5	0	0	4	ดี
ความพึงพอใจในภาพรวม	5	9	3	0	0	4.12	ดี
ค่าเฉลี่ยรวม						4.01	ดี

จากตารางที่ 18 พบร่วมกับค่าเฉลี่ยความพึงพอใจต่อระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) โดยภาพรวมอยู่ในเกณฑ์ประเมินที่ดี มีค่าเฉลี่ยเท่ากับ 4.01

ขอเสนอแนะ

สามารถสรุปความพึงพอใจและขอเสนอแนะของผู้ใช้ที่มีต่อ ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means) ได้ดังนี้

- ควรจะมีฟังก์ชันที่ให้เลือกจังหวัดว่าเราอยากลงของจังหวัดไหน เพื่อบางคนอยากรажานทำงานใกล้ๆบ้าน
- อยากรีบให้มีการเลือกโซนของภาคอย่างภาคเหนือหรือภาคอีสานแต่โดยรวมทำได้ดีแล้ว
- ต้องทำให้ผู้ใช้งาน ใช้งานได้ง่ายกว่านี้ และการค้นหาบ้างอย่างก็ไม่ตรงกับความต้องการที่ค้นหาเท่าไหร่

เอกสารอ้างอิง

- ตาเยะ, ช. (2022, 6 28). NLP คืออะไร . Retrieved from mindphp: <https://www.mindphp.com/คุณมีอะไร/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 (ซิสเต็ม อนาคต และ ดิจิทัล). (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). เทคนิคการคัดเลือกกลุ่มโหลดรายอาคารสำหรับรองรับแผนการติดตั้งระบบผลิตไฟฟ้าพลังงานแสงอาทิตย์บนหลังคาเพื่อเพิ่มค่าదารชนีประสิทธิภาพการใช้พลังงานไฟฟ้า Load Clustering Technique Application to PV Solar Rooftop Installation Planning for Improving Energy Efficiency, 134–148.
- ประณีญ พ่วงวิชา, อ. ท. (2019). ศึกษาการแบ่งกลุ่มพฤติกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ Clustering of Jewellery Purchasing Behaviour through Social Network, 213–224.
- จักรินทร์ สันติรัตนภักดี, ศ. น. (2021). การออกแบบและพัฒนากระบวนการจำแนกข้อมูลเรียนรถโดยสารสาธารณะเพื่อติดแท็กป้ายหากการให้บริการ, 77–90.

- วุฒิชัย, ว. (2013). การเปรียบเทียบวิธีการแบ่งแยกคำภาษาไทยด้วยโครงสร้างการเขียนกับโครงสร้างพยัญชนะ 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 cloudm: <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.glurgeek.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-8fbb36e68b0>

- 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? + สอนวิธีใช้งานเบื้องต้น. Retrieved from devahoy: <https://devahoy.com/blog/2015/08/getting-started-with-mongodb>

Natakorn, C. (2021, 6 3). FastAPI คืออะไร และการใช้งานเบื้องต้น. Retrieved from medium: <https://natakornch.medium.com/fastapi-คืออะไร-และการใช้งานเบื้องต้น-4f2d0fd91bcd>

TAeng Trirong, P. (2017, 12 25). Cross-Origin Resource Sharing (CORS) เป็นสิ่งที่ Web Developer ต้องควรรู้. Retrieved from medium: <https://medium.com/nellika/cors-เป็นสิ่งที่-web-developer-ต้องควรรู้-c906b1b47958>

Supalerk, P. (2020, 3 17). เมื่อสาย DATA อายากจะกิน Pizza (โดยใช้ Jaccard Similarity และ Cosine Similarity). Retrieved from medium: <https://medium.com/data-cafe-thailand/เมื่อสาย-data-อายากจะกิน-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). ลักษณะความสำคัญของข้อมูลความด้วยเทคนิคการประมวลผลทางภาษา เปื้องต้น: 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: แยกประเภทอีเมลล์ด้วยพลังฟอร์ซ. Retrieved from medium: <https://medium.com/mmp-li/nlp-natural-language-processing-ศาสตร์-ไม่-ใหม่-ศาสตร์แห่งเจ้า-แยกประเภทอีเมลล์ด้วยพลังฟอร์ซ-66b8bdff2e42>

រាជធានីភ្នំពេញ

គម្រោង

ການພັນວານ ກ
ຄູ່ມືອກາຮຕິດຕັ້ງ

การเติร์ยมข้อมูลและติดตั้งโปรแกรมสำหรับการพัฒนาระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเคลื่อน (K-Means)

1. การจัดกลุ่มข้อมูลด้วยเทคนิคทางปัญญาประดิษฐ์เคลื่อน (K-Means)

1.1 ทำการดาวน์โหลด Repository จาก https://github.com/slapexs/final_project ด้วยการใช้คำสั่ง `git clone https://github.com/slapexs/final_project` เพื่อดาวน์โหลดไฟล์ทั้งหมดมาไว้ในโฟเดอร์ของ

1.2 บันทึกไฟล์ข้อมูลบริษัทนามสกุล .csv ในไฟล์เอกสาร data.csv

1.3 เนื้อหา clustering.py ในโปรแกรม Text editor

ภาพที่ 58 โค้ดคำสั่งในไฟล์ clustering.py ใช้ในการจัดกลุ่มข้อมูล

1.4 ทำการกำหนดจำนวนของกลุ่มข้อมูลที่ต้องการลงในตัวแปร k

```

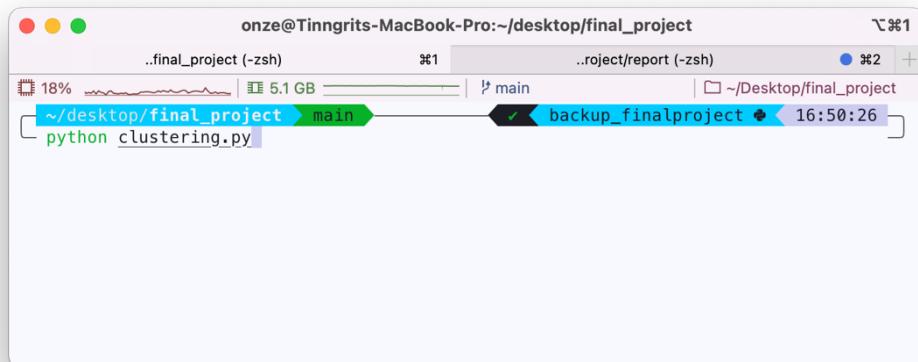
1 k = 7
2 kmeans = KMeans(n_clusters=k, random_state=1)
3 # Fit model
4 kmeans.fit(df_tfidf[['x_value', 'y_value']])
5 clusters = kmeans.labels_

```

ภาพที่ 59 แสดงกำหนดจำนวนของกลุ่มที่ต้องการลงในตัวแปร k และการจัดกลุ่มข้อมูล

จากภาพที่ 59 แสดงการจัดกลุ่มข้อมูลด้วยเคลื่อน (K-Means) โดยการที่กำหนดจำนวนของกลุ่มข้อมูลที่ต้องการลงในตัวแปร k จากนั้นทำการจัดกลุ่มข้อมูลด้วยฟังก์ชัน KMeans และคืนค่ากลับมาในตัวแปร kmeans และเก็บป้ายชื่อของกลุ่มลงในตัวแปร clusters

1.5 เรียกใช้ไฟล์เพื่อทำการจัดกลุ่มข้อมูล



ภาพที่ 60 แสดงการใช้งานคำสั่งจัดกลุ่มข้อมูลใน Terminal

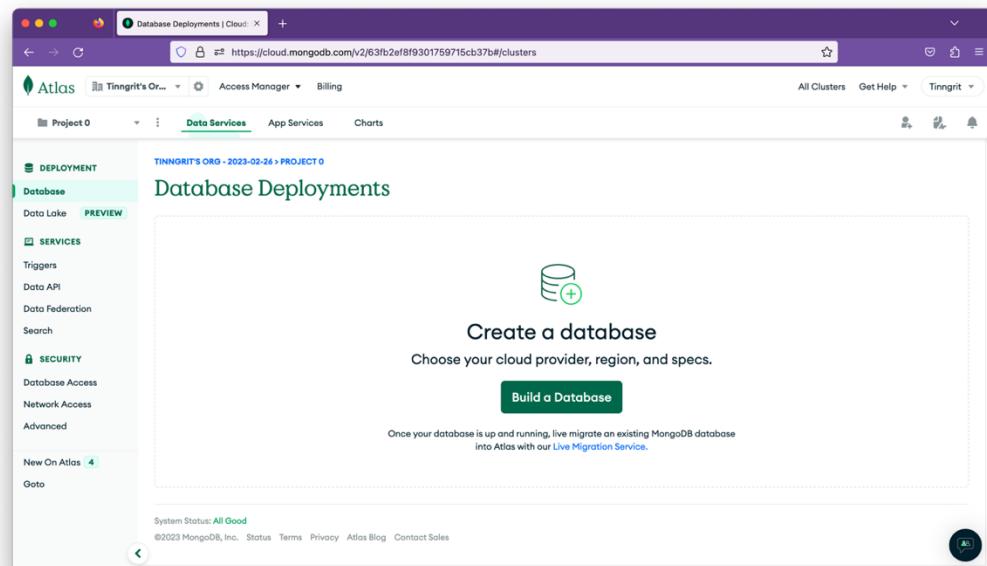
1.6 ได้ไฟล์ clustered_company.csv ในโฟลเดอร์ document ที่เป็นผลลัพธ์การจัดกลุ่มข้อมูล

h_company	Eng_compan	Type_business	Type_techno	Product	Type_innoval	Detail	Owner	Province_bas	Address	Phone_num	Email
บริษัท เทเลเมดิค้า	Telemedica Co	SW, Digital Se	เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	Platform	นางสาว กัญญา	กรุงเทพมหานคร	18,20 ซอยสุขุม 09-0004-000	support@oocci	
บริษัท ไดรฟ์เมต	Drivemate Co.	SW	เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	Platform	นางสาว ฐิติยา	กรุงเทพมหานคร	1168 ถนนเพชรบุรี	0-2026-3238	contact@drive
บริษัท นิวเอนเน	Venue Co., L	SW, Digital Se	เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	Platform	นาย พีพันธ์	กรุงเทพมหานคร	ชั้น 201 อาคาร C	0-2120-6013	admin@venue
บริษัท ฮับเคาร์	Haucar Co., I	SW	เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	Platform	นาย ภูญชัย	กรุงเทพมหานคร	128/403 ชั้น 3	0-2080-3960	customers@hi
บริษัท ชิปเปป	Shippop Co., I	SW, Digital Se	เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	Platform	นาย ภราวด์	กรุงเทพมหานคร	128/239-240	09-2905-335	contact@ship
บริษัท และท่าอากาศยาน	Airports Co., SW	Digital Se	เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	Platform	นาย ภราวด์	กรุงเทพมหานคร	128/217 หมู่บ้าน	06-3216-669	center@airpor
บริษัท ฟู้ดสตอรี่	Food Stories L	HW/Smart Dev			บริการ	Platform	นาย เอกพิพัฒน์	กรุงเทพมหานคร	29/217 หมู่บ้าน	06-3216-669	center@foodst
บริษัท บล็อก	Bulk Asia Co., SW				บริการ	Platform	นาย โน้ต	กรุงเทพมหานคร	ชั้น 14 อาคาร 2	0-2101-2851	support@bulk
บริษัท บล็อก อี	Bulk One Gro	SW			บริการ	Platform	นาย พศิษฐ์	กรุงเทพมหานคร	ชั้น 14 อาคาร 2	0-2101-2851	support@bulk
บริษัท เซลล์สุกี้	Sellsuki Co., L	SW			บริการ	Platform	นาย อนุพัฒ์	กรุงเทพมหานคร	10/39 ชั้น 3 ชั้น 3	0-2026-3250	hello@sellsuki
บริษัท โกลฟดิจิ	Golfdigg Co., I	SW	Digital Se		บริการ	Platform	นาย ภูญชัย	กรุงเทพมหานคร	5 ซอยวังน้ำ 4 ตึก 4	08-9333-100	hello@golfdig
บริษัท ไอ-แอป	I-App Creator	SW	Digital Se		บริการ	Platform	นาย เอกชัย	กรุงเทพมหานคร	23/34-35 อาคาร	08-8668-333	support@iapp
บริษัท แม่น้ำพลัง	Energy Respo	HW/Smart Dev	Artificial Intelli		บริการ	Platform	นาย ไบร์อัน	กรุงเทพมหานคร	408 อาคาร A	08-0264-614	contact@enne
บริษัท ทัช คอ	Touchcore Co.	SW	Digital Co	Chatbot	บริการ	Platform	นาย อัคราภรณ์	กรุงเทพมหานคร	69/5 อาคาร E	0-240-25626	care@onechat
บริษัท ดิจิทัล	Visual Edu Co.	SW			บริการ	Platform	นาย อินโนว์	กรุงเทพมหานคร	264/121 ชั้น 7	0-6419-094	hi@dexii.co
บริษัท ดิจิทัลฟาร์ม	Infoed Co., Lt	SW	Digital Co	Artificial Intelli	บริการ	Platform	นาย จิรภพ	กรุงเทพมหานคร	670 ถนนเพชรบุรี	08-8018-888	contact@info
บริษัท น้องสต็อก	OneStockholm	SW			บริการ	Platform	นาย เอกธิพัฒน์	กรุงเทพมหานคร	1-7 อาคารอสังหาริมทรัพย์	0-2026-3223	cs@onestock
บริษัท รีฟิน อินโน	Refinn Interna	SW	Artificial Intelli		บริการ	Platform	นาย มนัสวิช	กรุงเทพมหานคร	897 ซอยสุขุม	0-2080-3924	contact@refin
บริษัท ไพรีซ่า	Priceza Co., L	SW			บริการ	Shopppir	นาย อรุณรัตน์	กรุงเทพมหานคร	126/101 อาคาร 1	0-2860-6125	info@priceza
บริษัท สกิลแลนด์	Skilllane Educ	SW	Digital Se		บริการ	Platform	นาย ชีริษฐ์	กรุงเทพมหานคร	49/5 ถนนสุขุมวิท	08-1996-7156	info@SkillLan
บริษัท โลคัล ไลก์	Local Alike Co.	SW			บริการ	Platform	นาย สมศักดิ์	กรุงเทพมหานคร	12 ซอยอินทนนท์	0-2115-9861	info@localalik
บริษัท ลิ๊ก มี ช่า	Like Me Co., L	SW			บริการ	Platform	นาย ณัฐพัฒน์	กรุงเทพมหานคร	216 ซอยสุขุม	09-2264-549	hi@ammonne
บริษัท ลูกบี	Ookbee Co., L	SW	Digital Co	Artificial Intelli	องค์กร	กระบวนการผลิต	บริการ	กรุงเทพมหานคร	1104/207-209	0-2187-2701	member@ook
บริษัท จิตต้า	Jitta Dot Com	SW	Artificial Intelli		เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	กรุงเทพมหานคร	1111/127 หมู่บ้าน	0-2026-3707	wonderful@jitt
บริษัท ฟิกซี่ จั๊ก	Fixy Co., Ltd.	SW	Digital Se		เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	กรุงเทพมหานคร	599 ชั้น 3 ถนน	0-2165-0384	info@fixy.net
บริษัท โซเชียล จี	SocialMotion	SW	Digital Co		เทคโนโลยีสารสนเทศ	มีลักษณะ	บริการ	กรุงเทพมหานคร	อาคาร C ชั้น 4	06-2458-262	hello@socialg

ภาพที่ 61 แสดงไฟล์ clustered_company.csv

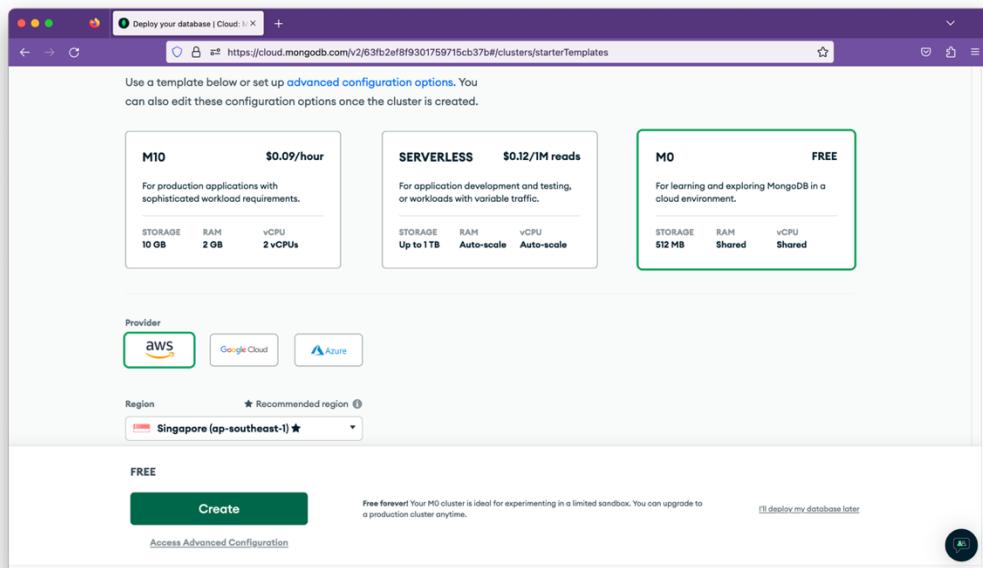
2. การสร้างคลัสเตอร์ (Cluster) ของฐานข้อมูลมองโกลดีบี (MongoDB) บนเว็บไซต์

- 2.1 การใช้งาน Cloud MongoDB เข้าเว็บไซต์ <https://www.mongodb.com> จากนั้นเข้าสู่ระบบ
- 2.2 สร้าง Cluster โดยการกดปุ่ม Build a Database



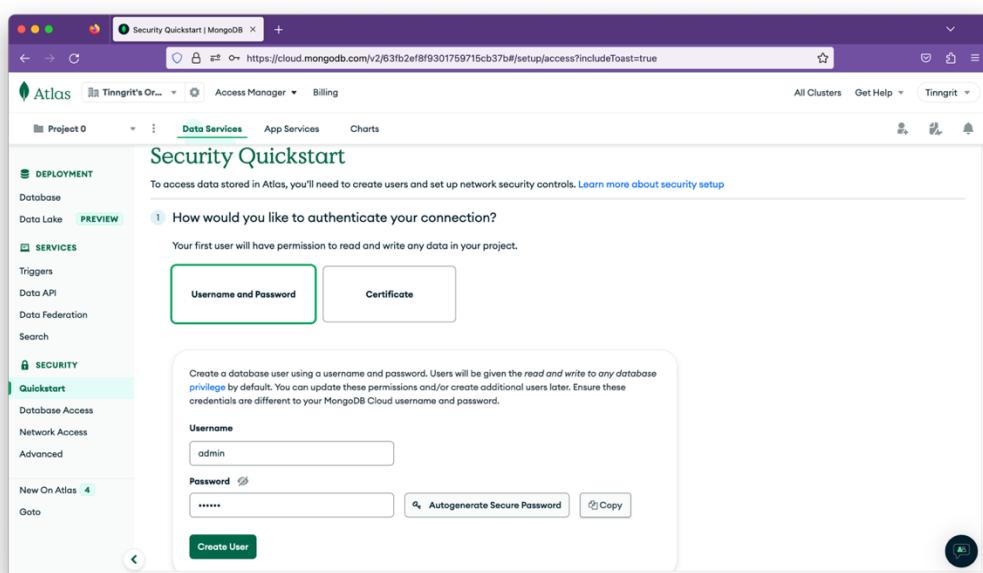
ภาพที่ 62 แสดงหน้าการจัดการ Cluster MongoDB

2.3 เลือกการตั้งค่าของ Cluster และตั้งชื่อ จากนั้นกดปุ่ม Create เพื่อสร้างฐานข้อมูลโดยที่เลือกผู้ให้บริการเป็น AWS และเลือกภูมิภาคเป็นประเทศไทย



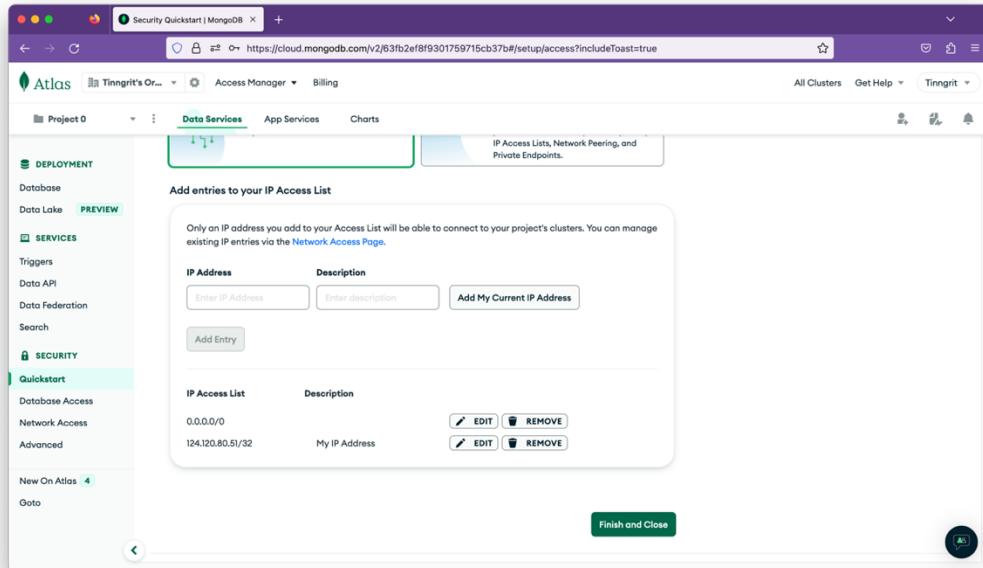
ภาพที่ 63 แสดงหน้าตั้งค่าและสร้าง Cluster MongoDB

2.4 สร้างบัญชีสำหรับใช้งานฐานข้อมูลตั้งค่า username และ password จากนั้นกดปุ่ม Create User



ภาพที่ 64 แสดงหน้าสร้างบัญชีสำหรับจัดการฐานข้อมูล

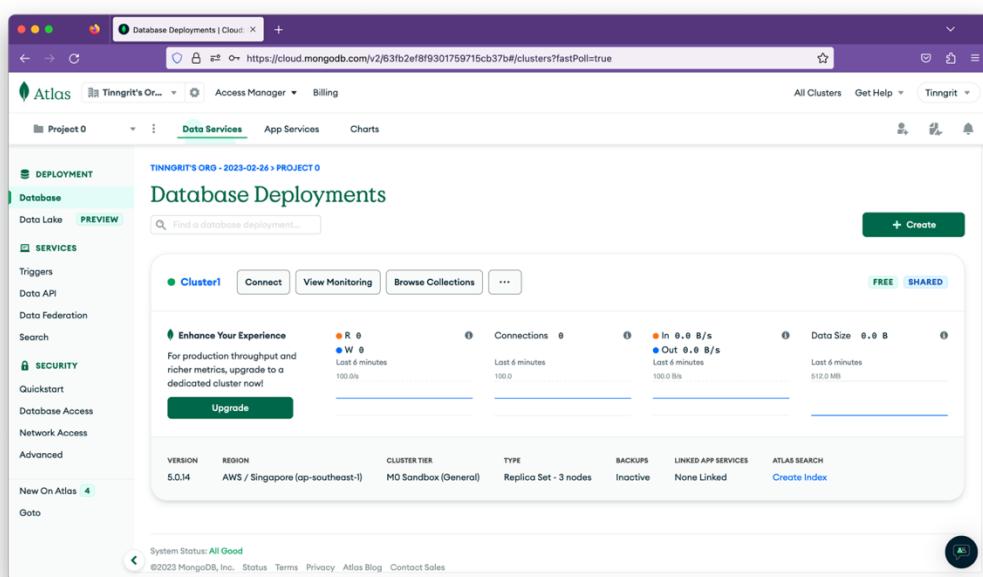
2.4 เพิ่มรายชื่อ IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้ (กรณีเป็น 0.0.0.0 หมายถึงทุก IP address สามารถเชื่อมต่อเข้ามาได้) จากนั้นกดปุ่ม Add Entry และปุ่ม Finish and Close



ภาพที่ 65 แสดงหน้าเพิ่ม IP address ที่สามารถเชื่อมต่อฐานข้อมูลได้

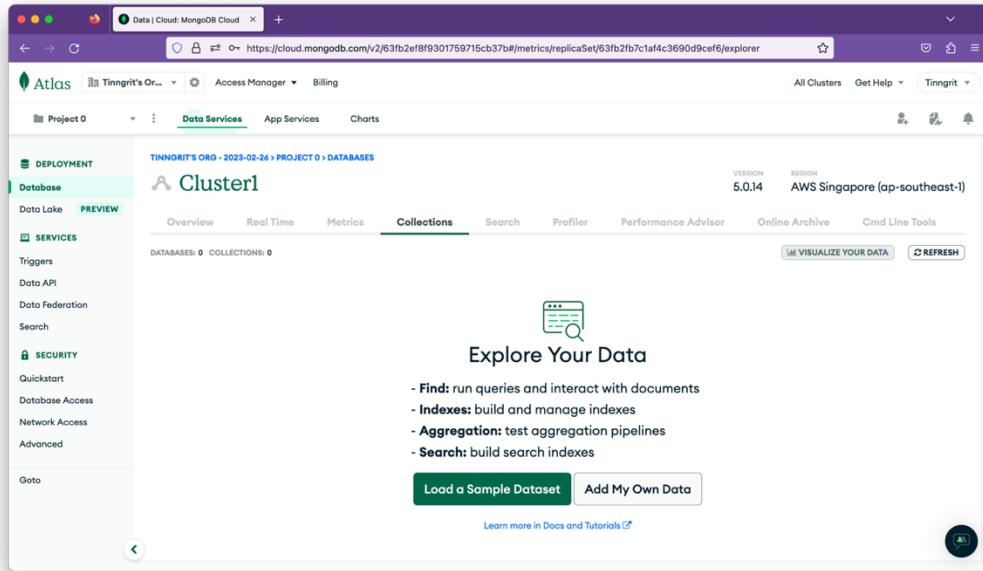
3. การสร้างฐานข้อมูลในโปรแกรม mongod (MongoDB)

3.1 สร้างฐานข้อมูลและ Collection กดที่ชื่อ Cluster



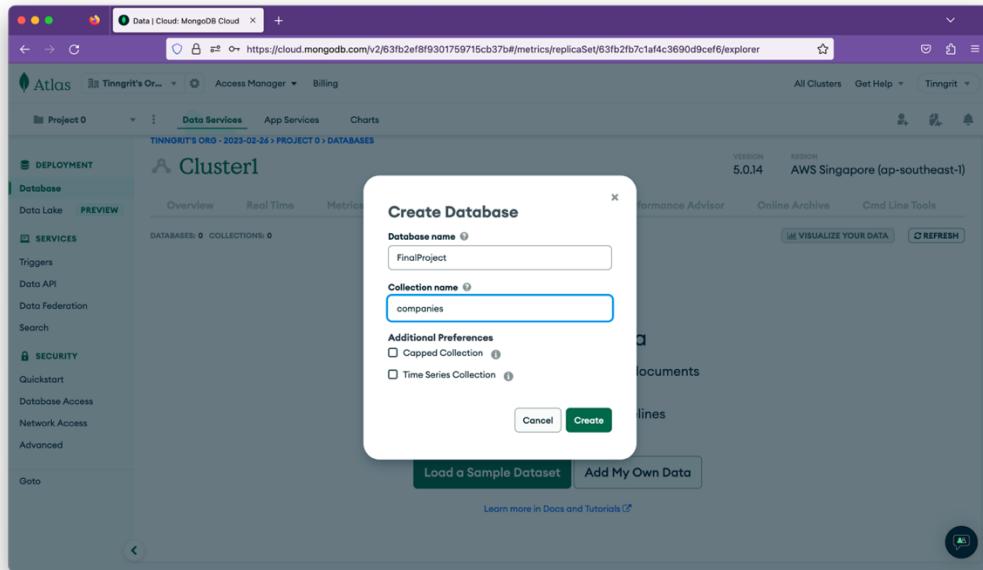
ภาพที่ 66 แสดงหน้าจัดการ Cluster MongoDB

3.2 กดที่ແນບເມື່ອ Collections ແລະກົດປຸ່ມ Add My Own Data ເພື່ອສ້າງຮູ້ານຂອ່ມລູ



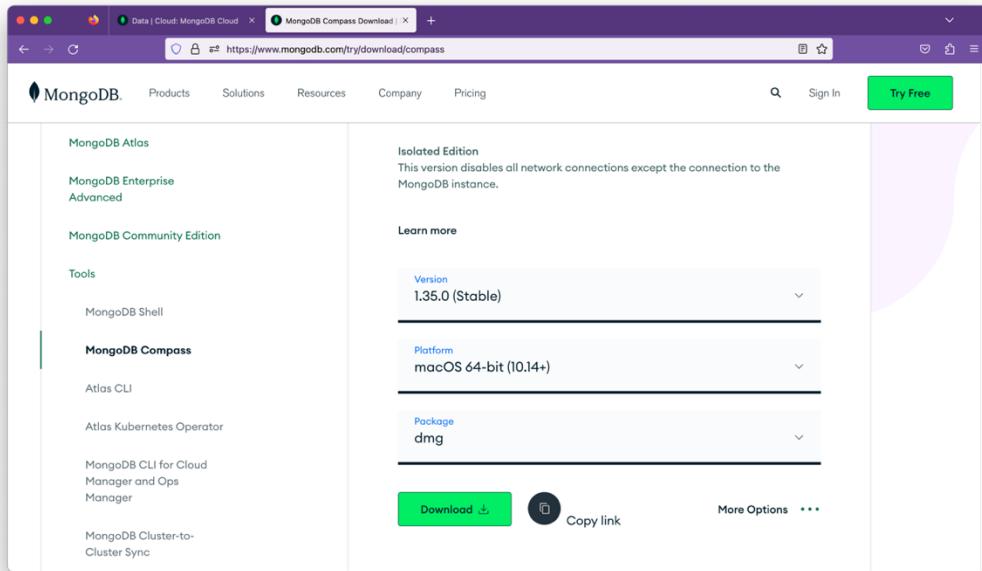
ກາພທີ 67 ແສດງກາຮສ້າງຮູ້ານຂອ່ມລູ MongoDB

3.3 ກໍາທັນດີຂໍ້ອຮູ້ານຂອ່ມລູແລະຂໍ້ອ Collection ທີ່ອຟູ້ໃນຮູ້ານຂອ່ມລູ



ກາພທີ 68 ແສດງໜ້າຕ່າງກາຮສ້າງຮູ້ານຂອ່ມລູແລະ Collection

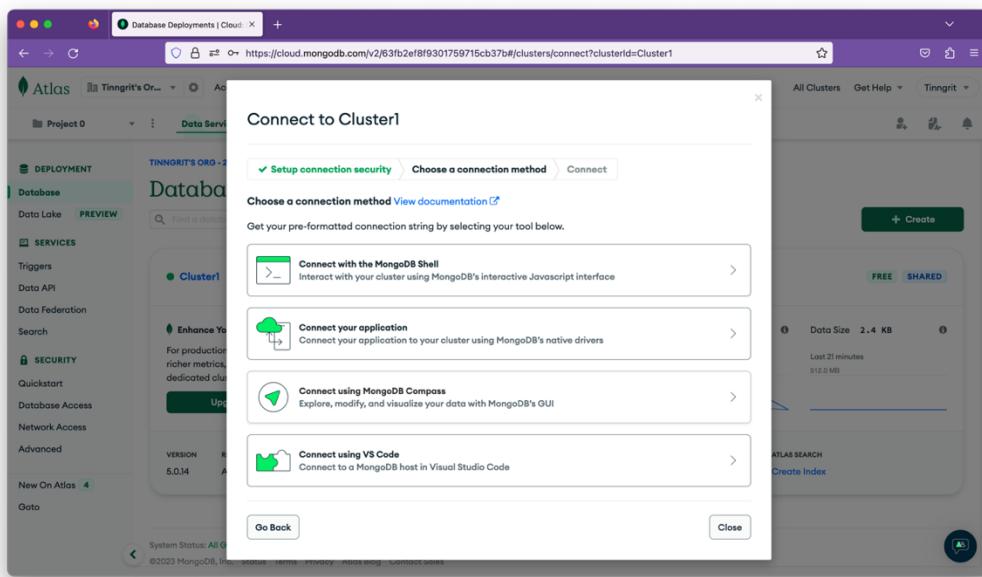
3.3 ດາວນີໂຫລດໂປຣແກຣມ MongoDB compass ທີ່ເວັບໄຊຕ່ອງ MongoDB ແລະເລືອກ Version Platform ແລະ Package ຕາມຮະບບປົງບັດກາຮທີ່ໃໝ່ແລະກົດປຸ່ມ Download



ภาพที่ 69 ตัวอย่างการเลือกตั้งค่าการดาวน์โหลดโปรแกรม MongoDB compass

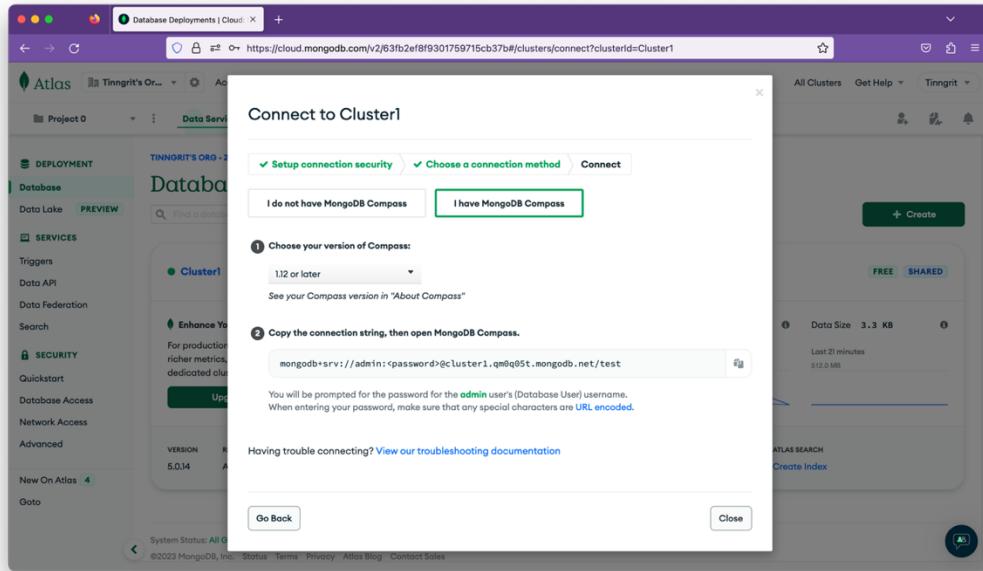
4. การนำข้อมูลเข้าสู่ฐานข้อมูล mongoddb (MongoDB)

4.1 เปิดเว็บไซต์หน้าจัดการ Cluster กดปุ่ม Connect และกดปุ่ม Connect using MongoDB Compass



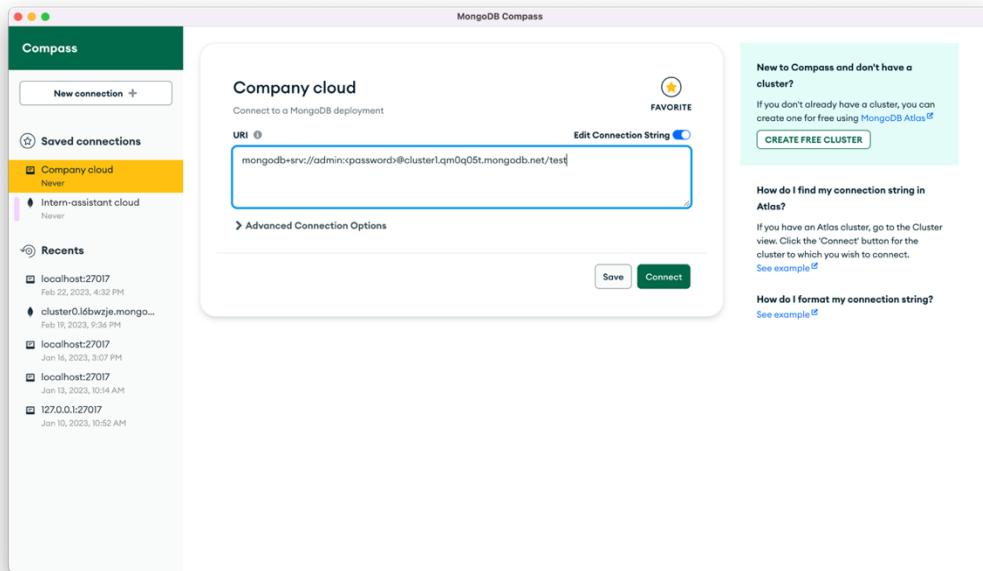
ภาพที่ 70 หน้าต่างเลือกเชื่อมต่อ กับ Cluster

4.2 กดปุ่ม I have MongoDB Compass และคัดลอก Connection string ให้ข้อ 2



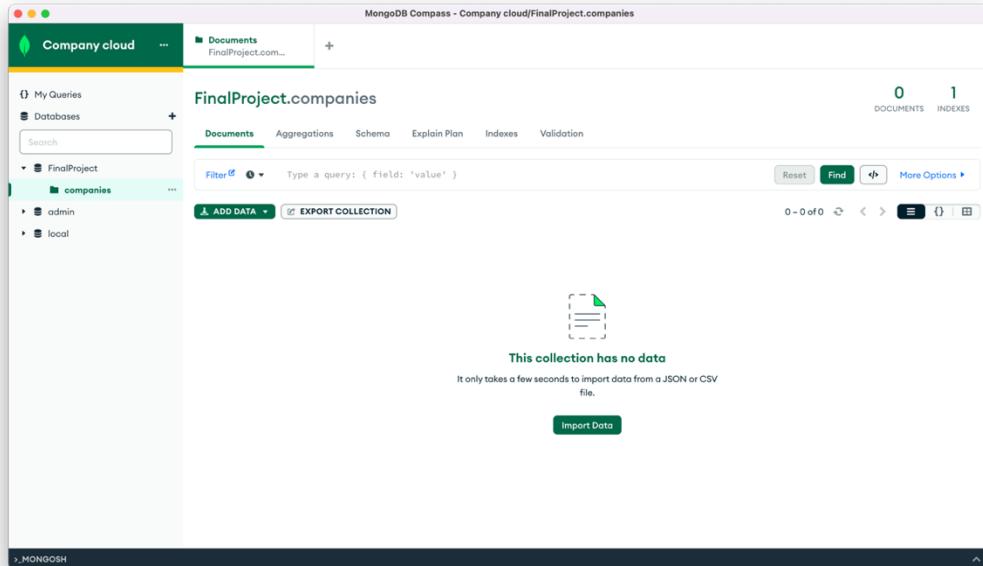
ภาพที่ 71 หน้าต่างข้อมูลการเชื่อมต่อ Cluster กับ MongoDB compass

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



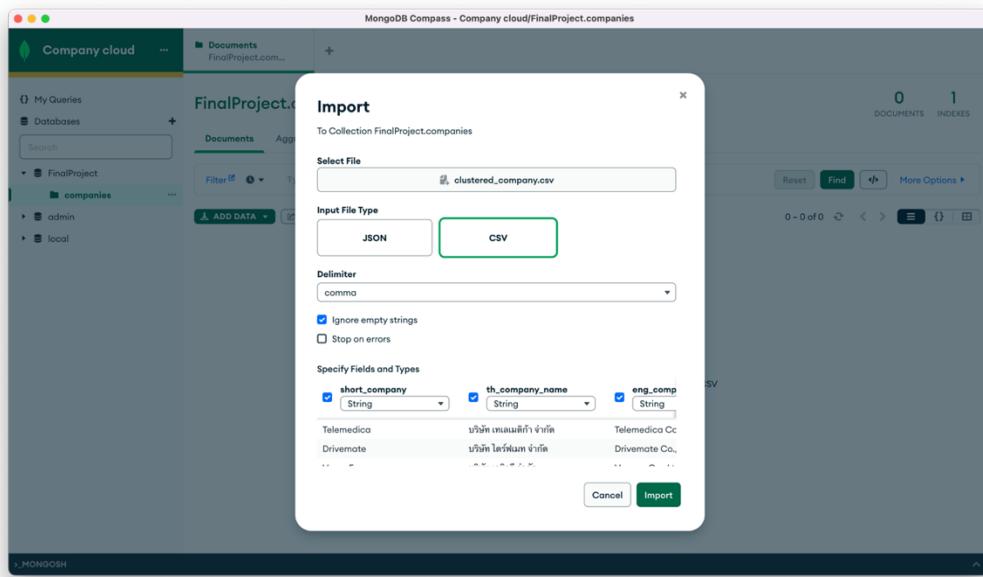
ภาพที่ 72 หน้าต่างโปรแกรม MongoDB compass สำหรับเชื่อมต่อ Cluster

4.4 เลือกฐานข้อมูลและ Collection ที่เมนูด้านซ้ายและกดปุ่ม Import Data



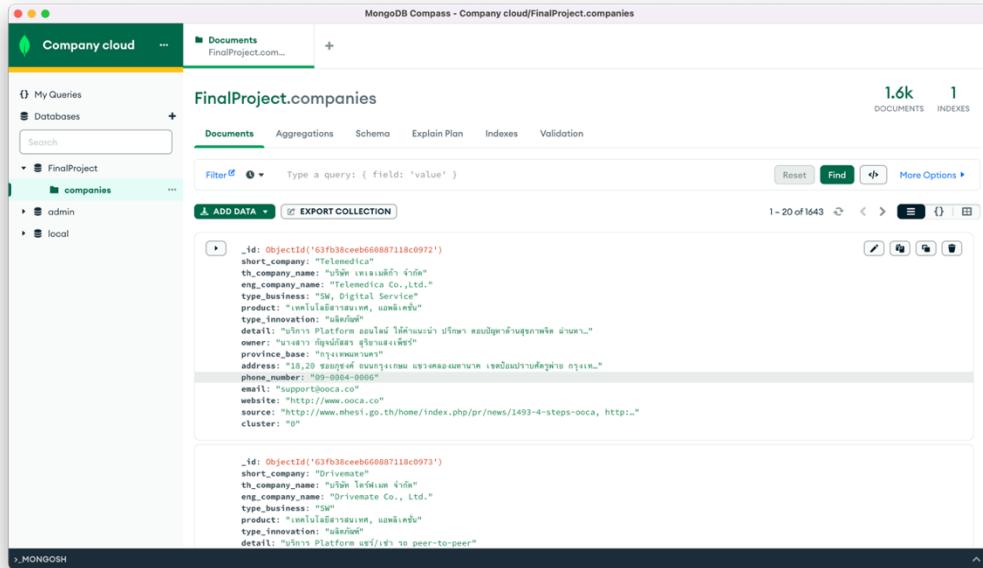
ภาพที่ 73 หน้าต่างโปรแกรมแสดงข้อมูลใน Collection

4.5 เลือกไฟล์ข้อมูลบริษัทที่จัดทำลุ่มแล้วกดปุ่ม CSV เพื่อ Import ข้อมูลแบบไฟล์นามสกุล csv และกดปุ่ม Import และกดปุ่ม Done เพื่อเสร็จลิ้นกระบวนการ



ภาพที่ 74 หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv

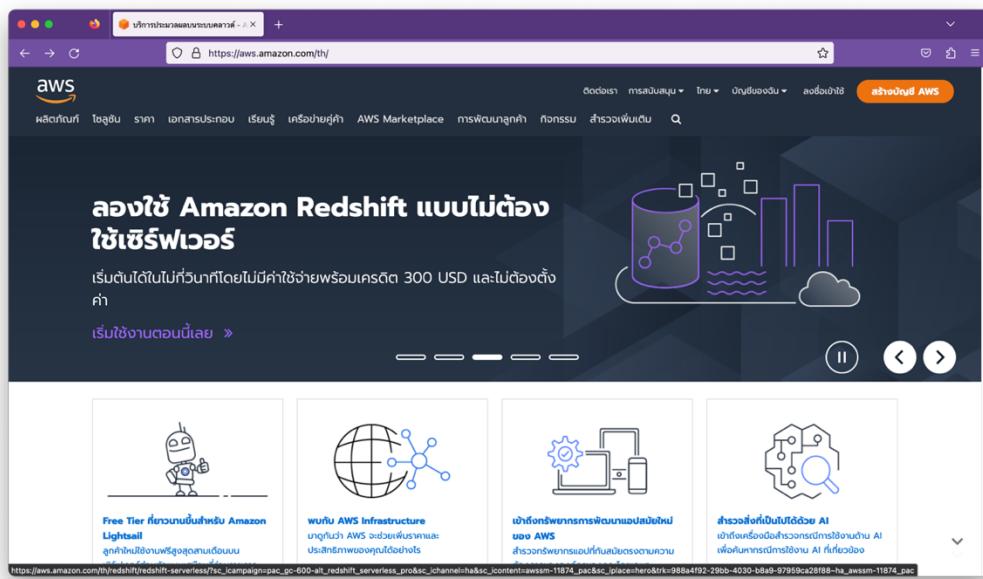
4.6 เมื่อ Import ข้อมูลสำเร็จจะได้ข้อมูลอยู่ใน Collection



ภาพที่ 75 หน้าต่างแสดงข้อมูลใน Collection ในโปรแกรม MongoDB compass

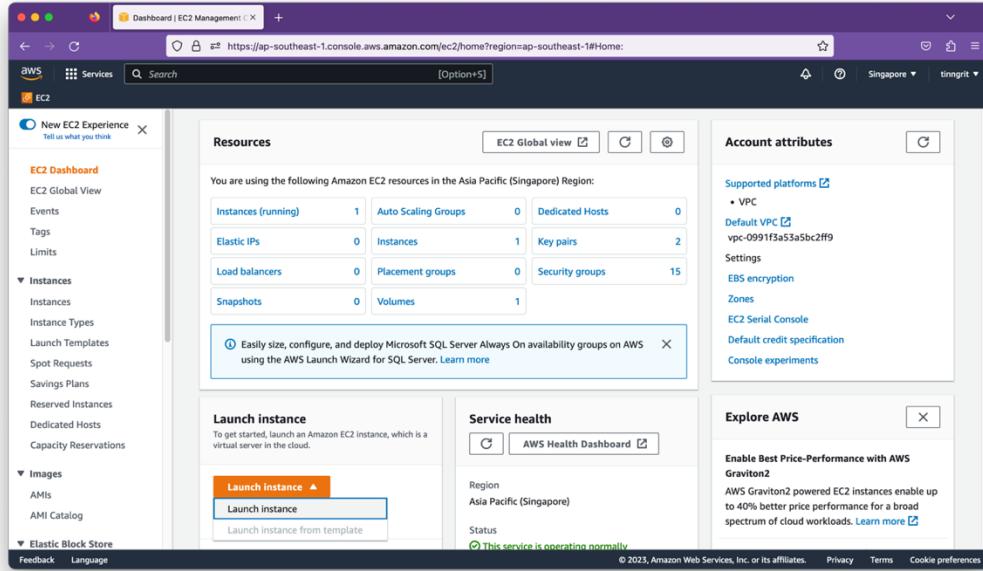
5. การใช้งาน Cloud computing ของ Amazon Web Services

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



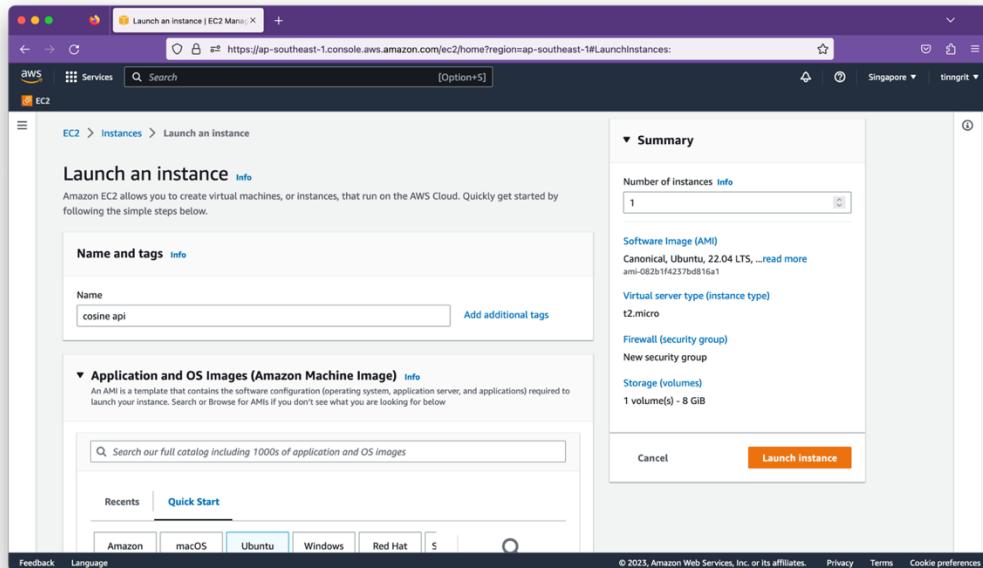
ภาพที่ 76 หน้าเว็บไซต์ Amazon Web Services

5.2 กดที่เมนู EC2 และกดปุ่ม Launch instance เพื่อสร้าง Instance ใหม่



ภาพที่ 77 หน้าแสดงการเลือกสร้าง Instance ใหม่

5.3 ตั้งค่าเครื่อง Instance ตามต้องการและกดปุ่ม Launch instance



ภาพที่ 78 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 79 ตัวอย่างการเชื่อมต่อเข้าไปยัง Instance

6. การติดตั้งและใช้งาน Web API สำหรับคำนวณค่าความคล้ายคลึง (Cosine similarity)

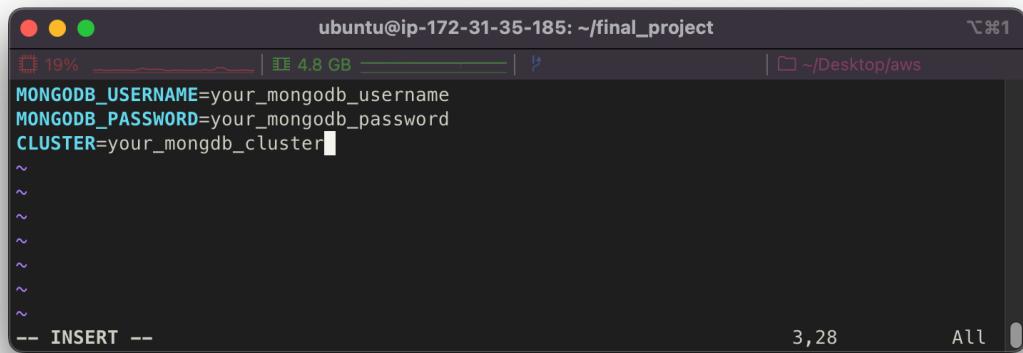
6.1 ทำการ Clone project จาก Github ที่ลิงค์ https://github.com/slapexs/final_project.git

```

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

ภาพที่ 80 การดาวน์โหลดโปรเจกจาก Github ด้วยคำสั่ง git clone

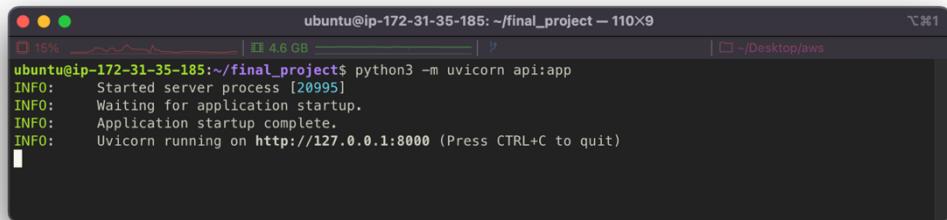
6.2 สร้างไฟล์ใหม่ในโฟลเดอร์ตั้งชื่อว่า .env และสร้างตัวแปรชื่อว่า MONGODB_USERNAME MONGODB_PASSWORD และ CLUSTER เพื่อใช้เก็บข้อมูลเชื่อมต่อฐานข้อมูล



```
ubuntu@ip-172-31-35-185: ~/final_project
MONGDB_USERNAME=your_mongodb_username
MONGDB_PASSWORD=your_mongodb_password
CLUSTER=your_mongodb_cluster
~
```

ภาพที่ 81 สร้างไฟล์ใหม่ชื่อ .env และสร้างตัวแปรเพื่อเก็บค่าเชื่อมต่อฐานข้อมูล

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



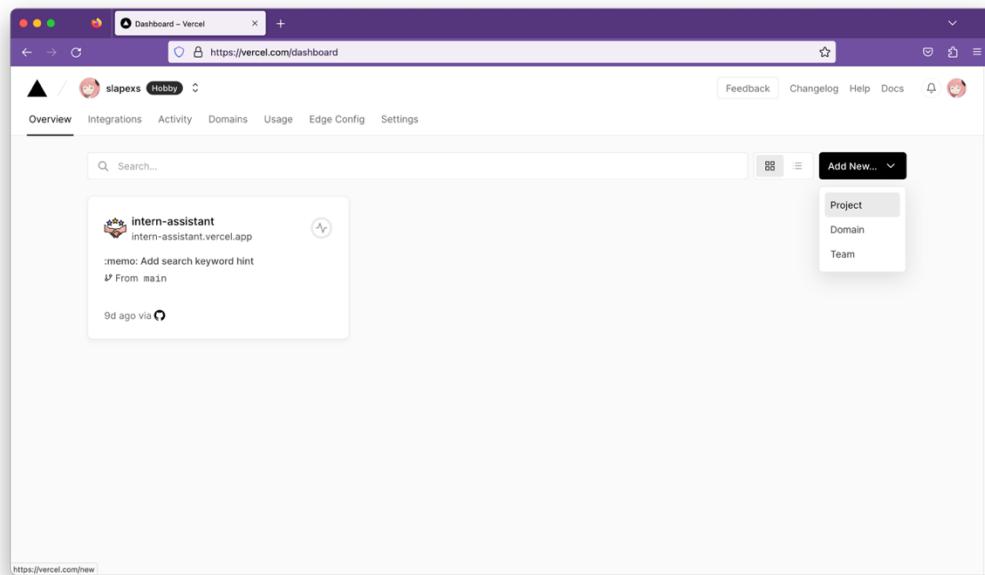
```
ubuntu@ip-172-31-35-185:~/final_project$ python3 -m uvicorn api:app
INFO:     Started server process [20995]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)
```

ภาพที่ 82 ตัวอย่างการเริ่มต้น Server Web API เพื่อคำนวณค่า Cosine similarity บน AWS

7. การติดตั้งและใช้งานเว็บแอปพลิเคชัน (Web application)

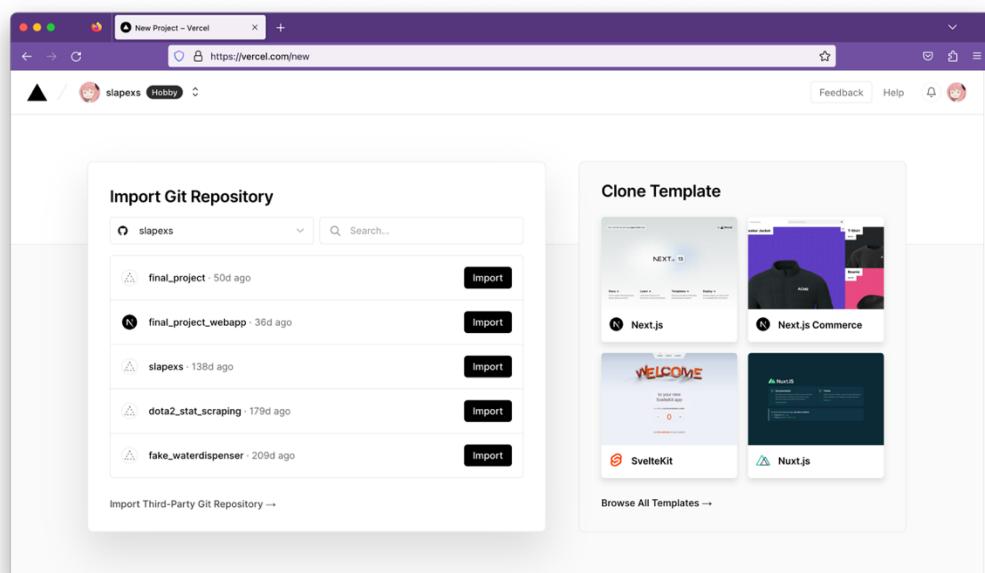
7.1 เข้าเว็บไซต์ <https://vercel.com> และเข้าสู่ระบบด้วยบัญชี Github

7.2 ทำการสร้างโปรเจคใหม่กดปุ่ม Add New และเลือก Project



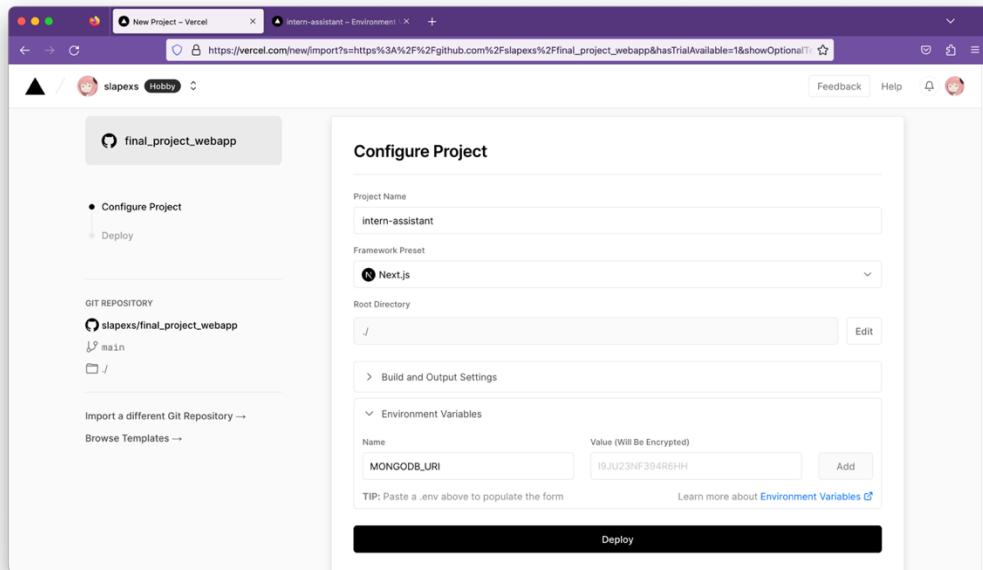
ภาพที่ 83 สร้างโปรเจคใหม่ใน Vercel

7.3 เลือก Repository ที่ต้องการจะ deploy และกดปุ่ม Import



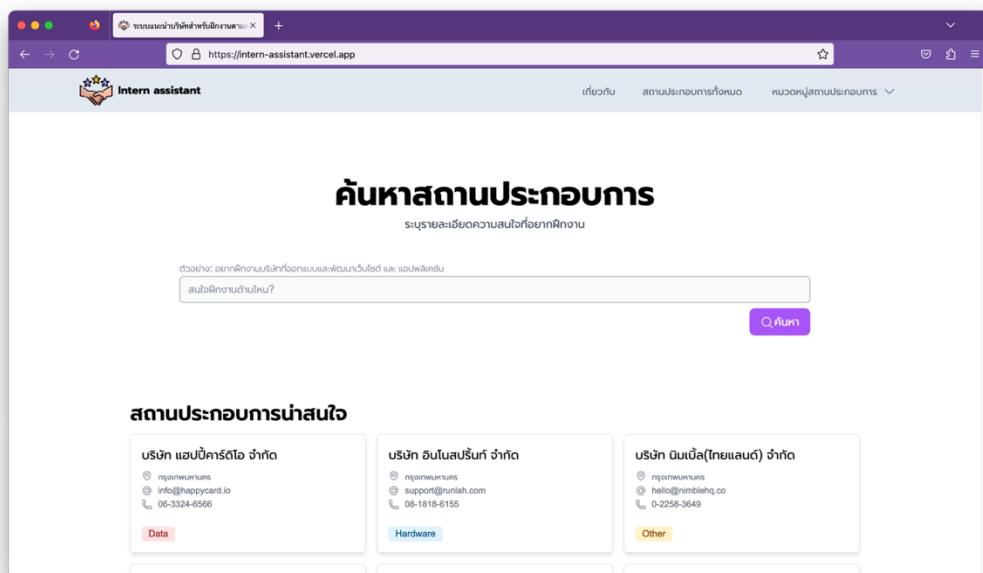
ภาพที่ 84 หน้าแสดงรายชื่อ Repository

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



ภาพที่ 85 หน้าการตั้งค่าโปรเจกตอน Deploy

7.5 ตัวอย่างหน้าเว็บไซต์เมื่อ Deploy เรียบร้อย



ภาพที่ 86 ตัวอย่างหน้าเว็บไซต์

រាជធានីភ្នំពេញ

កម្មីអការិច្ចាន

ภาคผนวก ข คู่มือการใช้งาน

ระบบแนะนำสถานที่ฝึกงานด้านคอมพิวเตอร์ ด้วยเทคโนโลยีการจัดกลุ่มเครื่อง (K-Means) สามารถแบ่งได้ 3 ส่วนดังนี้

1. ผู้ดูแลระบบ

1.1 การจัดกลุ่มข้อมูล

1 เปิดไฟล์ clustering.py ในโปรแกรม Text editor

2. ทำการกำหนดจำนวนของกลุ่มข้อมูลที่ตัวแปร k

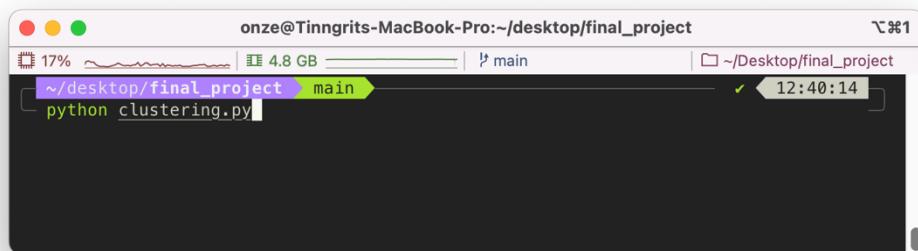
```

● ● ●
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 87 แสดงกำหนดจำนวนกลุ่มที่ตัวแปร k

3. ใช้คำสั่งใช้งานไฟล์เพื่อจัดกลุ่มข้อมูล



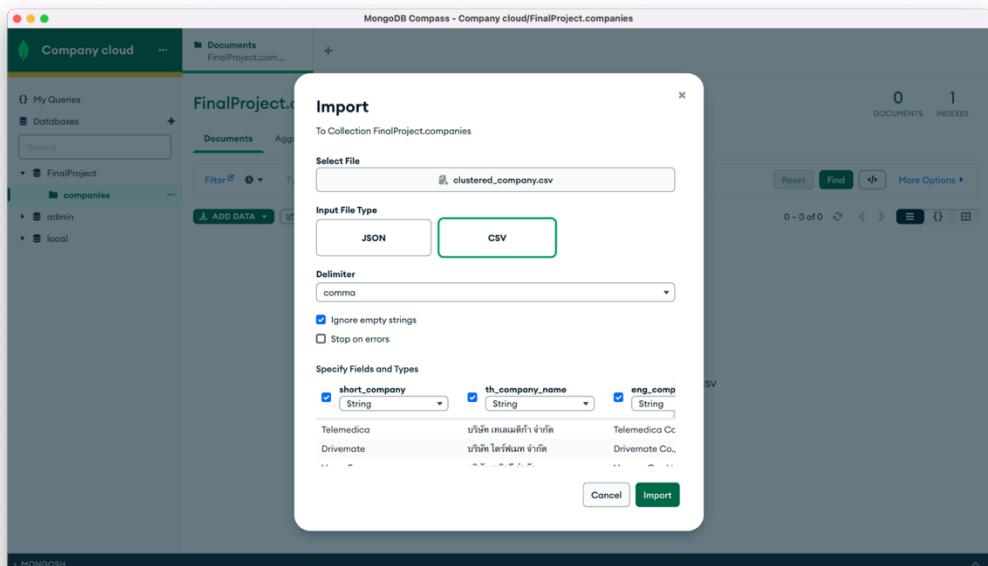
ภาพที่ 88 แสดงการใช้คำสั่งจัดกลุ่มข้อมูลใน Terminal

4. ได้ไฟล์ clustered_company.csv ในโฟลเดอร์ document ที่เป็นผลลัพธ์การจัดกลุ่มข้อมูล

h_company	Eng_company	Type_business	Type_techno	Product	Type_innovat	Detail	Owner	Province_ita	Address	Phone numb	Email
บริษัท เทเลเมดิค้า	Telematica Cc	SW, Digital Se	เทคโนโลยีสารสนเทศ	บริการ Platform	นวัตกรรม	กรุงเทพมหานคร 16,20 ชั้นชากูร 09-0004-000	support@occ				
บริษัท ไดรฟ์เมท	Drivemate Co.	SW	เทคโนโลยีสารสนเทศ	บริการ Platform	นวัตกรรม	กรุงเทพมหานคร 1168 ถนนเพชรบุรี 0-2026-3238 contact@driv					
บริษัท เวนิว จำกัด	Venue Co., L	SW, Digital Se	เทคโนโลยีสารสนเทศ	บริการ Platform	นวัตกรรม	กรุงเทพมหานคร ห้อง 201 อาคาร 2 0-2120-6013 admin@venu					
บริษัท ไฮปาร์ค	Haupcar Co., I	SW	เทคโนโลยีสารสนเทศ	บริการ Platform	นวัตกรรม	กรุงเทพมหานคร 128/403 ชั้น 3 0-2080-3960 customers@hi					
บริษัท ชิปปิ้ง	Shippop Co., I	SW, Digital Se	เทคโนโลยีสารสนเทศ	บริการ Platform	นวัตกรรม	กรุงเทพมหานคร 128/39-240 09-2905-335 contact@shpi					
บริษัท แอร์เพลส	Airportels Co., SW	Digital Se	เทคโนโลยีสารสนเทศ	บริการ Platform	นวัตกรรม	กรุงเทพมหานคร 29/217 หมู่บ้าน 06-3216-6691 center@airpor					
บริษัท ฟู้ดสตอรี่	Food Stories L	HWSmart Dev									
บริษัท บล็อก เชียร์	Bulk Asia Co., SW										
บริษัท บล็อก จำกัด	Bulk One Groi	SW									
บริษัท เซลสึคิ	Selsuki Co., L	SW									
บริษัท กอลดิจิ	Golddig Co., I	SW, Digital Se									
บริษัท ไอ-แอป	I-App Creator	SW, Digital Se									
บริษัท เอเนอร์จี้	Energy Respo	HWSmart Dev	Artificial Intelli								
บริษัท ทัฟ กอร์	Touchcore Co.	SW, Digital Co	Chatbot								
บริษัท วีดีโอ อีดี	Visual Edx Co.	SW									
บริษัท บีบี ไอที	Infobel Co., LT	SW, Digital Co	Artificial Intelli								
บริษัท วันเด็ก!	OneStockhouse	SW									
บริษัท รีฟิน อินโน	Refine Interna	SW	Artificial Intelli	เทคโนโลยีสารสนเทศ	แมตต์กันทร์						
บริษัท ไพริซ่า	Priciza Co., Li	SW									
บริษัท ศิลล์เอนดู	Skilllane Educ	SW, Digital Se									
บริษัท โลคัล ไลค์	Local Alike Co.	SW	เทคโนโลยีสารสนเทศ	แมตต์กันทร์							
บริษัท ไลค์ มี จำกัด	Like Me Co., L	SW									
บริษัท อุ๊คบี	Ookbee Co., L	SW, Digital Co	Artificial Intelli	อบรมการผลิต	บริการ เรียนรู้ดิจิทัล	กรุงเทพมหานคร 264/121 ถนนที่ 08-6419-094 hi@delex.co					
บริษัท จิตต้า	Jitta Dot Com	SW	Artificial Intelli	เทคโนโลยีสารสนเทศ	แมตต์กันทร์	กรุงเทพมหานคร 670 ถนนเพชรบุรี 08-8018-888 contact@inf					
บริษัท ฟิซี่	Fizzy Co., Ltd	SW, Digital Co	เทคโนโลยีสารสนเทศ	แมตต์กันทร์		บริการ เรียนรู้ดิจิทัล นวัตกรรม ดิจิทัล กรุงเทพมหานคร 1-7 ถนนสีลม 0-2026-3223 cs@onestockl					
บริษัท โซเชียลมูเตชั่น	SocialMotion	SW, Digital Co	เทคโนโลยีสารสนเทศ	แมตต์กันทร์		บริการ Shopppier ขาย อบรมดิจิทัล กรุงเทพมหานคร 897 ถนนสีลม 0-2080-3924 contact@refin					
						บริการ Shopppier ขาย อบรมดิจิทัล กรุงเทพมหานคร 126/101 ถนน 0-2860-6125 info@pricenza.j					
						บริการ Platform ขาย รีวิวดิจิทัล บนแอป 49/5 ถนนสีลม 08-1995-7161 info@skilllan					
						ลั่นทัญญิก้า	ขาย แมตต์กันทร์ กรุงเทพมหานคร 12 ซอยสีลม 0-2155-9861 info@localalki				
							บริการ เรียนรู้ดิจิทัล กรุงเทพมหานคร 216 ถนนสีลม 09-2264-549 hi@comonej				
							บริการ อบรมดิจิทัล นวัตกรรม ดิจิทัล กรุงเทพมหานคร 1104/207-209 0-2187-2701 member@ook				
							บริการ Platform ขาย ตรวจสอบ ดิจิทัล กรุงเทพมหานคร 1111/27 ถนน 0-2026-3707 wonderfullit				
							บริการ Platform ขาย ดิจิทัล เมือง ดิจิทัล กรุงเทพมหานคร 599 ชั้น 3 ถนน 0-2165-0384 info@fizzy.net				
							บริการ Platform ขาย ดิจิทัล เมือง ดิจิทัล กรุงเทพมหานคร 06-2458-262 hello@sociali				

ภาพที่ 89 แสดงไฟล์ clustered_company.csv

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



ภาพที่ 90 หน้าต่าง Import ข้อมูลนามสกุลไฟล์ csv

6. เมื่อ Import ข้อมูลสำเร็จจะได้ข้อมูลอยู่ใน Collection

The screenshot shows the MongoDB Compass interface with the database 'Company cloud' selected. Under the 'FinalProject' database, the 'companies' collection is highlighted. The interface shows two documents in the 'Documents' tab. The first document has the following fields:

- _id:** ObjectId("63fb38cced65887118c0972")
- short_company:** "Telemedica"
- th_company_name:** "บริษัท เทเลเมดิค้า จำกัด"
- eng_company_name:** "Telemedica Co.,Ltd."
- type_business:** "SM, Digital Service"
- product:** "imulusplatform, แพลตฟอร์ม"
- type_innovation:** "peer-to-peer"
- detail:** "แพลตฟอร์ม ให้เช่าบ้าน ปรึกษา และเชื่อมต่อผู้เช่ากับเจ้าของบ้าน."
- owner:** "นายวิวัฒน์ ศรีสุวรรณ"
- province_base:** "กรุงเทพมหานคร"
- address:** "18, 20 ถนนสุขุมวิท แขวงคลองเตยเหนือ เขตป้อมปราบศรีอุขุม กรุงเทพฯ"
- phone:** "+66 89 999 9999"
- email:** "support@ooca.co"
- website:** "http://www.ooca.co"
- source:** "http://www.nheeis.go.th/home/index.php/pr/news/1493-4-steps-ooca, http://..."
- cluster:** "0"

The second document is similar, with a different _id and slightly different product and detail descriptions.

ภาพที่ 91 ข้อมูลใน Collection ในโปรแกรม MongoDB compass

2. ผู้ใช้งาน

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

The screenshot shows the 'Intern assistant' website. At the top, there is a search bar with the placeholder 'ค้นหา' (Search) and a button labeled 'ค้นหา' (Search). Below the search bar, the text 'ระบุรายละเอียดความสนใจของคุณก่อน' (Specify your interests before proceeding) is displayed. The main area is titled 'ค้นหาสถานประกอบการ' (Search for business). It shows a grid of six cards representing different business categories:

- บริษัท ไทย ดอก คอบ จำกัด** (IT)
- บริษัท ชัยลพธ จำกัด** (Software)
- บริษัท บันเทิงเบลจ จำกัด** (Data)
- บริษัท ชลธรอาเจร์ จำกัด**
- บริษัท แอลปีซี จำกัด**
- บริษัท เอ็กซากอน แอลส จำกัด**

ภาพที่ 92 หน้าเว็บไซต์ intern-assistant.vercel.app

2.2 เมื่อเจอรายชื่อบริษัทผู้ใช้สามารถดูข้อมูลบริษัทได้ตามต้องการ

บริษัท เกจเมเด็ก จำกัด	บริษัท ไอ-แอป ครีเอชั่น จำกัด
บริษัท อินไฟฟ์ จำกัด	บริษัท สวีฟต์ ไบเนอร์ิกส์ จำกัด
บริษัท แม็ค เทค อินเตอร์ จำกัด	บริษัท โนว์โซลูชั่น จำกัด
บริษัท ออโต้เพียร์ จำกัด	บริษัท โลเคิล ออลิก จำกัด
บริษัท มีเดียโซล จำกัด	บริษัท เบลลู๊ก จำกัด

ภาพที่ 93 หน้าแสดงผลลัพธ์เมื่อค้นหาบริษัท

2.3 ผู้ใช้สามารถดูข้อมูลบริษัทเพื่อประกอบการตัดสินใจในการเลือกสถานประกอบการเพื่อ ฝึกงานได้

ภาพที่ 94 หน้าแสดงข้อมูลบริษัท

3. Web API

เป็นเว็บ API ที่ทำหน้าที่ให้ข้อมูลที่ต้องการและมีหน้าที่ในการคำนวณค่าความคล้ายคลึง (Cosine similarity) ข้อมูลที่สามารถศึกษาไปยังคำขอได้มีดังนี้

3.1 ข้อมูลบริษัททั้งหมดในฐานข้อมูล

```

1 {
2   "short_company": "Telemedica",
3   "th_company_name": "บริษัท เทเลเมดิก้า จำกัด",
4   "eng_company_name": "Telemedica Co.,Ltd.",
5   "type_business": "SM, Digital Service",
6   "product": "แพลตฟอร์มสุขภาพ แพทย์ ผู้ป่วย",
7   "type_innovation": "แพลตฟอร์ม",
8   "detail": "แพลตฟอร์มทางการแพทย์ที่เชื่อมต่อผู้ป่วยกับแพทย์ ให้บริการ ดูแลสุขภาพ ระยะยาว",
9   "owner": "บริษัท เทเลเมดิก้า จำกัด"
10 }
  
```

ภาพที่ 95 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/allcompanies>

จากภาพที่ 95 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/allcompanies> ด้วย เมธอด GET เพื่อขอข้อมูลบริษัททั้งหมดในฐานข้อมูลด้วยโปรแกรม Postman

3.2 ข้อมูลบริษัทที่อยู่ในกลุ่มที่กำหนด

```

1 {
2   "short_company": "SOTTLAVU",
3   "th_company_name": "บริษัท สอนลั่ว จำกัด",
4   "eng_company_name": "Sottlavu Co., Ltd.",
5   "type_business": "SM, SW",
6   "detail": "บริษัทซอฟต์แวร์ (Online Platform - Web/Mobile Application) สำหรับธุรกิจ (Business Management Software) ภายใต้ชื่อ LauvPOS (Lauv's Point of Sale System)",
7   "owner": "ประเทศไทย บริษัท สอนลั่ว จำกัด",
8   "province_base": "สุพรรณบุรี",
9   "address": "หมู่บ้าน ไอล์ฟ บี บ้านเลขที่ 59/19 หมู่ 2 ตำบลราษฎร์ อำเภอราษฎร์ จังหวัดสุพรรณบุรี"
10 }
  
```

ภาพที่ 96 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/company/1>

จากภาพที่ 96 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/company/1> ด้วย เมธอด GET เพื่อขอข้อมูลบริษัทที่อยู่ในกลุ่มที่ 1 ทั้งหมดในฐานข้อมูลด้วยโปรแกรม Postman

3.3 ข้อมูลรายการรหัสของกลุ่ม (Cluster ID) ทั้งหมดที่มีในฐานข้อมูล

```

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

```

ภาพที่ 97 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/cluster>

จากภาพที่ 97 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/cluster> ด้วยเมธอด GET เพื่อขอข้อมูลรายการรหัสของกลุ่มข้อมูล (Cluster ID) ทั้งหมด ในฐานข้อมูลด้วยโปรแกรม Postman

3.4 การส่งคำเพื่อคำนวณค่าความคล้ายคลึง (Cosine similarity)

1. ส่งกลับข้อมูลมาเป็นผลลัพธ์ค่าความคล้ายคลึงและรหัสกลุ่ม (Cluster ID)

```

1
2   "cosine_similarity": [
3     0.032153948733223255,
4     0.0,
5     0.1150986421365805,
6     0.18426464644709992,
7     0.023957046458011518,
8     0.3040314451234591,
9     0.03532800669540177
10    ],
11   "max_cosine_similarity": 0.3040331445123691,
12   "cluster": 5
13

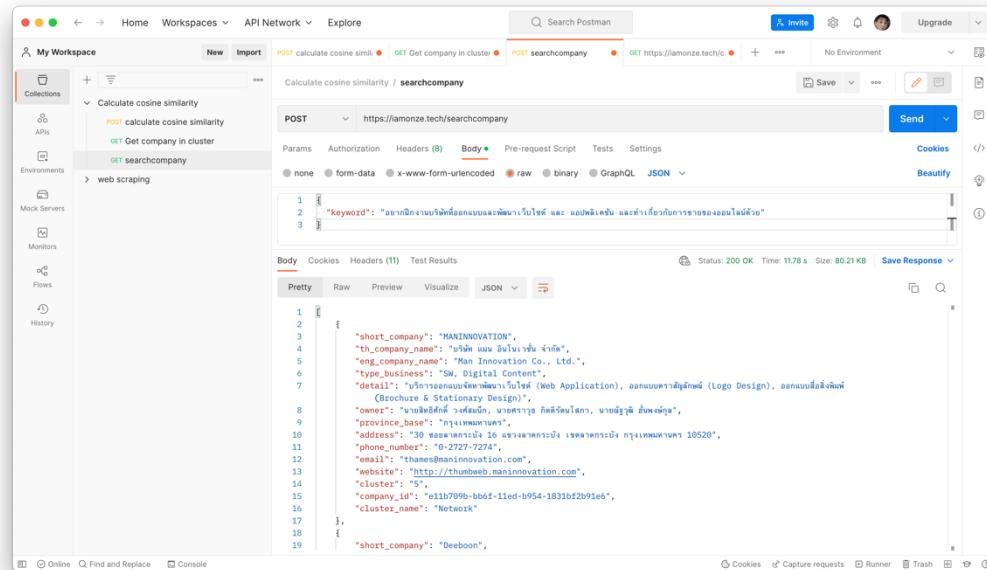
```

ภาพที่ 98 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/search>

จากภาพที่ 98 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/search> และแนบข้อมูลรายละเอียดความสนใจของผู้ใช้เป็นรายเพื่อคำนวณค่าความคล้ายคลึง (Cosine similarity) ด้วยเมธอด

POST และผลลัพธ์ที่คืนค่ากลับมาจะเป็นค่าความคล้ายคลึงของแต่ละกลุ่ม และรหัสกลุ่มข้อมูลที่มีความคล้ายคลึงมากที่สุด

2. ส่งกลับข้อมูลมาเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่มีความคล้ายมากที่สุด



```

POST https://iamonze.tech/searchcompany
{
  "Keyword": "รายงานการวิเคราะห์อุปกรณ์และแพรวาโนร์"
}
  
```

The screenshot shows the Postman interface with a collection named 'Calculate cosine similarity'. A POST request is selected with the URL 'https://iamonze.tech/searchcompany'. The 'Body' tab is active, showing a JSON object with a single key 'Keyword' and a value of 'รายงานการวิเคราะห์อุปกรณ์และแพรวาโนร์'. The response tab shows a status of 200 OK with a response body containing company details.

```

{
  "short_company": "MANINNOVATION",
  "th_company": "บริษัท มานิโนเวชั่น จำกัด",
  "eng_company": "Manin Innovation Co., Ltd.",
  "type_business": "SM, Digital Content",
  "detail": "บริษัทมานิโนเวชั่น จำกัด (Web Application), ออกแบบตราсти๊ก (Logo Design), ออกแบบเอกสาร (Brochure & Stationery Design)",
  "owner": "นายวิวัฒน์ วงศ์สมบูรณ์, นายวรรธน์ กิตติพันธุ์ไกร, นายปรัชญ์ อัมพุช",
  "province_base": "กรุงเทพมหานคร",
  "address": "38 ถนนพหลโยธิน 16 แขวงลาดยาว เขตคลองเตย กรุงเทพมหานคร 10520",
  "phone_number": "0-2727-7274",
  "email": "manin@maninnovation.com",
  "website": "http://thumweb.maninnovation.com",
  "cluster": "5",
  "company_id": "e11b799b-bb6f-11ed-b954-1831bf2b91e6",
  "cluster_name": "Network"
},
{
  "short_company": "Deeboon",
}
  
```

ภาพที่ 99 ตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/searchcompany>

จากภาพที่ 99 แสดงตัวอย่างการส่งคำขอไปยัง <https://iamonze.tech/searchcompany> และแนบข้อมูลรายละเอียดความสนใจของผู้ใช้ไปด้วยเพื่อคำนวณความคล้ายคลึง (Cosine similarity) มีลักษณะคล้ายกับการทำงานในภาพที่ 98 แต่ผลลัพธ์ที่คืนค่ากลับมาจะเป็นข้อมูลบริษัทที่อยู่ในกลุ่มที่มีความคล้ายมากที่สุด

ประวัติผู้ศึกษา



ชื่อ-นามสกุล : นายทินกร ลิงห์แก้ว

รหัสนักศึกษา : 64342205007-7

วันเดือนปีเกิด : 22 สิงหาคม พ.ศ. 2541

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

E-mail : tinngrit@outlook.com

ประวัติการศึกษา

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

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

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

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