



DSI@TU

Data Science & Innovation
Thammasat University

รายงานฉบับสมบูรณ์

การวิเคราะห์ความสอดคล้องของรายวิชาเพื่อเป็นแนวทางใน
การเทียบโอนหน่วยกิตในระบบธนาคารหน่วยกิตของ
มหาวิทยาลัยธรรมศาสตร์

โดย

- | | |
|------------------------------|-----------------------|
| 1. นายปณณวิช ศิลปเสรีฐ | เลขทะเบียน 6524651038 |
| 2. นางสาวกชพร สิทธิไพศาล | เลขทะเบียน 6524651145 |
| 3. นางสาวณัฏชา คงแก้ว | เลขทะเบียน 6524651236 |
| 4. นางสาวลภัสรดา ตรีกตรองกิจ | เลขทะเบียน 6524651376 |
| 5. นางสาวหนึ่งฤทัย ทองนัด | เลขทะเบียน 6524651442 |

อาจารย์ที่ปรึกษาโครงการ: ผู้ช่วยศาสตราจารย์ ดร. รัชฎา คงคะจันทร์

รายงานนี้เป็นส่วนหนึ่งของรายวิชา วช314 โครงการด้านการวิเคราะห์ข้อมูลสำหรับธุรกิจ
(DSI314 Business Analytics Capstone Project)

ภาคการศึกษา 1/2567

สารบัญ

สารบัญ	หน้า
สารบัญรูปภาพ	ข
สารบัญตาราง	ค
กิตติกรรมประกาศ	ง
บทสรุปผู้บริหาร (Executive Summary)	จ
บทที่ 1 บทนำ	1
1.1. ความสำคัญ / ที่มาของปัญหา	1
1.2. วัตถุประสงค์ของโครงการ	1
1.3. ขอบเขตของโครงการ	1
1.4. ประโยชน์ที่ได้รับของโครงการ	2
บทที่ 2 ทบทวนวรรณกรรม	3
2.1. ความรู้พื้นฐาน	3
2.2. งานที่เกี่ยวข้อง	3
2.3. ตารางเปรียบเทียบโครงการกับงานอื่นที่มีอยู่ในปัจจุบัน	6
บทที่ 3 การดำเนินงานโครงการ	7
3.1. การรวบรวมข้อมูล (Data Collection/ Data Acquisition)	7
3.2. การเตรียมข้อมูลเบื้องต้น (Data Preprocessing / Data Cleaning)	9
3.3. การวิเคราะห์ภาพรวมข้อมูล (Exploratory Data Analysis / Data Visualization)	12
3.4. การสร้างแบบจำลอง (Model Building)	14
3.5. การนำแบบจำลองไปใช้งาน (Model Deployment)	16
บทที่ 4 ผลการดำเนินงานโครงการ	19
2.4. DataFrame	19
2.5. Website	20
บทที่ 5 ความเชื่อมโยงกับวิชาต่าง ๆ ในโมดูล	22
5.1. วข.310 การสำรวจและการเตรียมข้อมูล	22
5.2. วข.311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล	23
5.3. วข.312 ระบบธุรกิจอัจฉริยะ	24
5.4. วข.313 การวิเคราะห์การตลาด	25
บทที่ 6 บทสรุป	26
บรรณานุกรม	27

สารบัญรูปภาพ

รูป	หน้า
รูปที่ 3-1 เว็บไซต์สำนักงานทะเบียนนักศึกษา มหาวิทยาลัยธรรมศาสตร์	8
รูปที่ 3-2 หน้าเว็บไซต์เมื่อทำการเลือกรายละเอียดต่าง ๆ	8
รูปที่ 3-3 ตัวอย่างหน้าเว็บข้อมูลรายวิชา	9
รูปที่ 3-4 Scraping Data DataFrame	9
รูปที่ 3-5 Convert Data Type	9
รูปที่ 3-6 ชูดข้อมูลใน Google Sheet	10
รูปที่ 3-7 lowercase	10
รูปที่ 3-8 Course Description Standard	10
รูปที่ 3-9 Drop NaN and Duplicate	10
รูปที่ 3-10 Course Code Standard	11
รูปที่ 3-11 RegEx	11
รูปที่ 3-12 DataFrame	12
รูปที่ 3-13 pie chart แสดงจำนวนรายวิชาทั้งหมดของ 5 คณะ (คิดเป็น %)	13
รูปที่ 3-14 กราฟแท่งแบบต่อแสดงค่าความถี่ของคำที่พบมากที่สุดในการอธิบายรายวิชาของแต่ละคณะ	13
รูปที่ 3-15 SBERT Model	14
รูปที่ 3-16 Embeded Column	15
รูปที่ 3-17 Cosine Similarity	16
รูปที่ 3-18 Home page	17
รูปที่ 3-19 Pair Course Page	18
รูปที่ 4-1 Final DataFrame	19
รูปที่ 4-2 Sidebar	20
รูปที่ 4-3 Course Description compare	21
รูปที่ 5-1 รูปแสดงชุดข้อมูลรายวิชา ก่อนทำ Data Preprocessing	22
รูปที่ 5-2 ชุดข้อมูลรายวิชา หลังทำ Data Preprocessing	23
รูปที่ 5-3 pie chart แสดงจำนวนรายวิชาทั้งหมดของ 5 คณะ (คิดเป็น %)	24
รูปที่ 5-4 กราฟแท่งแบบต่อแสดงค่าความถี่ของคำที่พบมากที่สุดในการอธิบายรายวิชาของแต่ละคณะ	24

สารบัญตาราง

ตาราง

หน้า

ตารางที่ 2-1 ตารางเปรียบเทียบโครงการกับงานอื่นที่มีอยู่ในปัจจุบัน

6

กิตติกรรมประกาศ

โครงการฉบับนี้เป็นส่วนหนึ่งของวิชา วช314 โครงการด้านการวิเคราะห์ข้อมูลสำหรับธุรกิจ โดยมีวัตถุประสงค์เพื่อให้คณะผู้จัดทำได้ฝึกการประยุกต์ใช้ความเชื่อมโยงกับวิชาต่าง ๆ ในโมดูล ได้แก่ วช310 การสำรวจและการเตรียมข้อมูล, วช311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล, วช312 ระบบธุรกิจอัจฉริยะ และ วช313 การวิเคราะห์การตลาด

ทั้งนี้โครงการสามารถสำเร็จลุล่วงได้อย่างสมบูรณ์จาก ผู้ช่วยศาสตราจารย์ ดร. รัชฎา คงคะจันทร์ ที่ได้สละเวลาอันมีค่าแก่คณะผู้จัดทำเพื่อให้คำปรึกษาและแนะนำ ตลอดจนตรวจทานแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่เป็นอย่างยิ่ง จนโครงการฉบับนี้สำเร็จลุล่วงได้ด้วยดี คณะผู้จัดทำขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

คณะผู้จัดทำ

กิตติกรรมประกาศ

โครงการฉบับนี้เป็นส่วนหนึ่งของวิชา วช314 โครงการด้านการวิเคราะห์ข้อมูลสำหรับธุรกิจ โดยมีวัตถุประสงค์เพื่อให้คณะผู้จัดทำได้ฝึกการประยุกต์ใช้ความเชื่อมโยงกับวิชาต่าง ๆ ในโมดูล ได้แก่ วช310 การสำรวจและการเตรียมข้อมูล, วช311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล, วช312 ระบบธุรกิจอัจฉริยะ และ วช313 การวิเคราะห์การตลาด

ทั้งนี้โครงการสามารถสำเร็จลุล่วงได้อย่างสมบูรณ์จาก ผู้ช่วยศาสตราจารย์ ดร. รัชฎา คงคะจันทร์ ที่ได้สละเวลาอันมีค่าแก่คณะผู้จัดทำเพื่อให้คำปรึกษาและแนะนำ ตลอดจนตรวจทานแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่เป็นอย่างยิ่ง จนโครงการฉบับนี้สำเร็จลุล่วงได้ด้วยดี คณะผู้จัดทำขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

คณะผู้จัดทำ

บทสรุปผู้บริหาร

ธนาคารหน่วยกิต (Credit Bank) ช่วยให้มีโอกาสในการเข้าถึงการศึกษา เพื่อพัฒนาทักษะ และส่งเสริมการเรียนรู้ตลอดชีวิต โดยเป็นระบบทะเบียนสะสมหน่วยกิตและกลไกในการเทียบโอนความรู้หรือสมรรถนะที่ได้จากการศึกษาในระบบ การศึกษานอกระบบ การศึกษาตามอัธยาศัย และจากประสบการณ์บุคคล สำหรับเก็บสะสมไว้ในธนาคารหน่วยกิตแห่งชาติ ซึ่งเชื่อมโยงกับธนาคารหน่วยกิตของสถาบันอุดมศึกษา ทำให้ผู้เรียนสามารถลงทะเบียนเรียนและสะสมหน่วยกิตได้ตลอดชีวิต โดยไม่มีเงื่อนไขของระยะเวลาในการสะสมและระยะเวลาในการศึกษา และสามารถเทียบโอนหน่วยกิตระหว่างสถาบันอุดมศึกษาได้ ซึ่งในปัจจุบันสถาบันระดับอุดมศึกษาหลายแห่งได้มีการกำหนดนโยบายร่วมกันเพื่อส่งเสริมการเรียนรู้ตลอดชีวิตและการพัฒนากำลังคนภาคอุตสาหกรรมในรูปแบบธนาคารหน่วยกิต

ในระดับอุดมศึกษา ทางสถาบันต่าง ๆ สามารถเทียบโอนรายวิชาภายในสถาบันหรือต่างสถาบันได้ ทางมหาวิทยาลัยธรรมศาสตร์เองก็สามารถเทียบโอนได้เช่นกัน โดยในการเทียบโอนรายวิชาภายในมหาวิทยาลัยธรรมศาสตร์ ปัจจุบันทำได้เฉพาะบางรายวิชาเท่านั้น

ทางคณะผู้จัดทำจึงสนใจและมีความต้องการศึกษาและวิเคราะห์ข้อมูลคำอธิบายรายวิชาจากรายวิชาที่เปิดสอนภายในมหาวิทยาลัยธรรมศาสตร์ โดยมุ่งเน้นไปยังหลักสูตรการศึกษาระดับปริญญาตรี ภาคการศึกษาที่ 1/2566 และ 2/2566 และมุ่งเน้นคณะที่เกี่ยวข้องกับเทคโนโลยีจำนวน 5 คณะ ได้แก่ คณะวิศวกรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, วิทยาลัยสหวิทยาการ, วิทยาลัยนวัตกรรม และสถาบันเทคโนโลยีนานาชาติสิรินธร (SIIT) เพื่อนำข้อมูลที่ได้จากการวิเคราะห์ความคล้ายคลึงกันของเนื้อหาการเรียนการสอนในรายวิชาต่าง ๆ ภายในคณะที่เกี่ยวข้อง ไปใช้เพื่อช่วยสนับสนุนการเทียบโอนรายวิชาและหน่วยกิตภายในมหาวิทยาลัยธรรมศาสตร์ และเพื่อนำข้อมูลที่ได้จากการวิเคราะห์มาช่วยสนับสนุนและเป็นแนวทางสำหรับการทำธนาคารหน่วยกิต (Credit Bank) ในอนาคต

กระบวนการวิเคราะห์หาความสอดคล้องของรายวิชาต่าง ๆ ที่มีความคล้ายคลึงกัน มีการเก็บรวบรวมมาจากเว็บไซต์ https://web.reg.tu.ac.th/registrar/class_info.asp?lang=en โดยขั้นตอนที่ 1 วิธีการ Web Scraping และเก็บข้อมูลรหัสรายวิชา (Course Code), คำอธิบายรายวิชา (Course Description), รายวิชาบังคับก่อน (Disciplinary Course) และรายวิชาต่อเนื่อง (Consecutive Course) จาก 5 คณะภายในมหาวิทยาลัย ธรรมศาสตร์ที่ได้กล่าวมาข้างต้น โดยหลังจากนั้นขั้นตอนที่ 2 ได้ทำการทำ Data Processing ใช้เทคนิคต่าง ๆ เช่น Strip, RegEx และ Lowercase เพื่อทำให้เกิดมาตรฐานของข้อมูล และขั้นตอนที่ 3 Exploratory Data Analysis (EDA) ด้วยการทำกราฟแท่งและกราฟต่อเนื่อง ในการแสดงสัดส่วนแต่ละรายวิชาของแต่ละคณะ และความถี่ของคำที่พบบ่อยในแต่ละคณะ ส่วนในขั้นตอนที่ 4 คือการนำ vector ที่เป็นตัวแทนของแต่ละ Course Description มาหาความคล้ายคลึงกัน โดยใช้ Cosine Simirality และเลือกใช้โมเดล 3 ตัว คือ all-MiniLM-L6-v2 , paraphrase-mpnet-base-v2 และ all-mpnet-base-v2 แล้วนำผลลัพธ์มาบันทึกใน DataFrame และขั้นตอนที่ 5 ขึ้นสุดท้าย นำผลลัพธ์สุดท้ายมาแสดงให้เห็นและเข้าใจง่ายในรูปแบบเว็บไซต์ จากการทำ streamlit framework สำหรับผู้ที่ต้องการใช้ข้อมูลในการเทียบโอนรายวิชา อีกทั้งยังเป็นแนวทางสำหรับธนาคารหน่วยกิตของมหาวิทยาลัยธรรมศาสตร์

บทที่ 1 บทนำ

1.1. ความสำคัญ / ที่มาของปัญหา

การสะสมหน่วยการเรียนรู้ หรือธนาคารหน่วยกิต (Credit Bank) เป็นการดำเนินการเพื่อเปิดโอกาสให้ผู้เรียนและประชาชนได้นำผลการเรียนรู้และประสบการณ์การทำงานหรืออาชีพทั้งในระบบ นอกระบบ และตามอัธยาศัย มาทำการรับรองและเทียบโอนกันได้ โดยมุ่งหวังให้ผู้เรียนและประชาชนมีโอกาสเชื่อมโยงการเรียนรู้หรือประสบการณ์การทำงาน สามารถพัฒนาตนเองมากยิ่งขึ้น อีกทั้งยังส่งเสริมการเรียนรู้ตลอดชีวิต (life long learning) และสร้างวัฒนธรรมการเรียนรู้ตลอดชีวิตให้มหาวิทยาลัย ซึ่งผู้เรียนที่ออกจากระบบมหาวิทยาลัยไปแล้ว สามารถเข้ามาเรียนใหม่ได้ตลอดเวลา โดยลดข้อจำกัดบางอย่าง เช่น การลงทะเบียนเรียนใหม่ การไม่จำกัดอายุของผู้เข้าเรียน

ในการศึกษาระดับอุดมศึกษา สถาบันระดับอุดมศึกษาได้มีการเปิดให้สามารถเทียบโอนรายวิชา ระหว่างสถาบันหรือระหว่างคณะสาขาภายในสถาบันเดียวกันได้ เพื่อนักศึกษาจะได้ไม่ต้องเรียนวิชานั้นซ้ำ รวมถึงมหาวิทยาลัยธรรมศาสตร์ นักศึกษาก็สามารถทำการเทียบโอนรายวิชาได้เช่นกัน แต่ในปัจจุบันการเทียบโอนรายวิชาในมหาวิทยาลัยธรรมศาสตร์สามารถทำได้เฉพาะบางวิชาเท่านั้น ดังนั้นทางคณะผู้จัดทำจึงต้องการที่จะศึกษาและวิเคราะห์รายวิชาที่นอกเหนือจากรายวิชาที่สามารถเทียบโอนได้ในปัจจุบัน ว่ามีโอกาสในการเทียบโอนกับรายวิชาใด ๆ ได้บ้างหรือไม่ โดยจะใช้ข้อมูลในส่วนของคำอธิบายรายวิชาจากรายวิชาที่เปิดสอนภายในมหาวิทยาลัยธรรมศาสตร์ ซึ่งจะมุ่งเน้นไปยังคณะที่เกี่ยวข้องกับเทคโนโลยีทั้งหมด 5 คณะด้วยกัน ได้แก่ คณะวิศวกรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, วิทยาลัยสหวิทยาการ, วิทยาลัยนวัตกรรมการ และสถาบันเทคโนโลยีนานาชาติสิรินธร (SIIT) เพื่อนำข้อมูลที่ได้จากการวิเคราะห์ไปใช้ในการช่วยสนับสนุนการเทียบโอนรายวิชาและหน่วยกิตภายในมหาวิทยาลัยธรรมศาสตร์ รวมไปถึงการใช้เป็นแนวทางในการสร้างธนาคารหน่วยกิต (Credit Bank) สำหรับมหาวิทยาลัยธรรมศาสตร์ในอนาคต

1.2. วัตถุประสงค์ของโครงการ

- 1.2.1. สามารถนำผลลัพธ์ที่ได้จากการวิเคราะห์ความสอดคล้องของรายวิชาต่าง ๆ ในคณะที่เกี่ยวข้องไปใช้เพื่อช่วยสนับสนุนการตัดสินใจในเรื่องของรายวิชาที่สามารถเทียบโอนหน่วยกิตกันได้ภายในมหาวิทยาลัยธรรมศาสตร์
- 1.2.2. เพื่อนำผลลัพธ์ที่ได้จากการวิเคราะห์มาช่วยสนับสนุนและเป็นแนวทางสำหรับการทำธนาคารหน่วยกิต (Credit Bank) ของมหาวิทยาลัยธรรมศาสตร์ในอนาคต

1.3. ขอบเขตของโครงการ

การรวบรวมข้อมูลรหัสรายวิชา (Course Code) คำอธิบายรายวิชา (Course Description) รายวิชาบังคับก่อน (Disciplinary Course) และ รายวิชาต่อเนื่อง (Consecutive Course) ที่เปิดสอนในหลักสูตรการศึกษาระดับปริญญาตรี ภาคการศึกษาที่ 1/2566 และ 2/2566 ภายในคณะที่เกี่ยวข้องกับเทคโนโลยี 5

คณะด้วยกัน ได้แก่ คณะวิศวกรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, วิทยาลัยสหวิทยาการ, วิทยาลัยนวัตกรรม และสถาบันเทคโนโลยีนานาชาติสิรินธร(SIIT) มหาวิทยาลัยธรรมศาสตร์ โดยใช้วิธีการดึงข้อมูลจากเว็บไซต์ TUREG โดยใช้ Selenium และใช้ SBERT model ซึ่งเป็น Transformers-based model ในกระบวนการวิเคราะห์คำอธิบายรายวิชาพร้อมแปลงคำเป็น vector และใช้เทคนิค Cosine Similarity ในการเทียบหาความสอดคล้องของรายวิชาต่าง ๆ เพื่อเป็นประโยชน์แก่ผู้ที่มีส่วนเกี่ยวข้องหรือผู้ที่สนใจ ทั้งในด้านการเทียบโอนรายวิชาและหน่วยกิตภายในมหาวิทยาลัยธรรมศาสตร์ และด้านการทำธนาคารหน่วยกิต (Credit Bank) ในอนาคต

1.4. ประโยชน์ที่ได้รับของโครงการ

จากการวิเคราะห์ข้อมูลเพื่อหาความสอดคล้องของรายวิชาต่าง ๆ ทำให้เกิดประโยชน์มากมายหลายด้าน ด้านมหาวิทยาลัยธรรมศาสตร์ การที่สามารถนำผลลัพธ์ของการวิเคราะห์ไปสนับสนุนการเทียบโอนรายวิชาและหน่วยกิต พร้อมทั้งใช้เป็นแนวทางในการสร้างระบบธนาคารหน่วยกิตของมหาวิทยาลัยได้ เพื่อมุ่งสู่การเป็นมหาวิทยาลัยที่ส่งเสริมการศึกษาตลอดชีวิตให้แก่ผู้เรียน ด้านนักศึกษา การลดค่าใช้จ่ายของนักศึกษาในการลงทะเบียนเรียนซ้ำในรายวิชาที่มีเนื้อหาคล้ายคลึงกัน, การลดเวลาในการเรียนรายวิชาที่เนื้อหามีความซ้ำซ้อน เพื่อเพิ่มเวลาให้นักศึกษาในการหาความรู้ใหม่เพิ่มเติม

บทที่ 2 ทบทวนวรรณกรรม

2.1. ความรู้พื้นฐาน

2.1.1. ประกาศกระทรวงศึกษาธิการ เรื่อง แนวทางการดำเนินงานธนาคารหน่วยกิตระดับการศึกษา ขั้นพื้นฐานและการอาชีวศึกษา พ.ศ. 2567

2.1.2. ความรู้เบื้องต้นเกี่ยวกับโมเดลที่ใช้

- Transformer เป็นโมเดลที่พัฒนาโดยบริษัท Google ถูกนำเสนอครั้งแรกในปี 2017 ในบทความงานวิจัยชื่อ Attention is All You Need โมเดล Transformer ทำงานโดยการรับ input sequence ที่มีลักษณะเป็นลำดับที่มีความยาว เช่นข้อความต่าง ๆ จากนั้น โมเดลจะหาความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ที่ได้รับเข้ามาและใช้ความสัมพันธ์นั้นในการสร้าง output sequence ออกมาเป็นคำตอบ โดยคำตอบจะมีลักษณะเป็นลำดับที่มีความยาวเช่นกัน

- BERT (Bidirectional Encoder Representations from Transformers) เป็นโมเดลประมวลผลภาษาธรรมชาติ (NLP) เป็นโมเดล Machine Learning (ML) ที่ถูกพัฒนาโดย Google ซึ่งถูกออกแบบมาเพื่อช่วยให้คอมพิวเตอร์เข้าใจความหมายของภาษาที่มีความกำกวมในข้อความได้ โดยใช้ข้อความรอบข้างเพื่อสร้างบริบท (Context)

- SBERT (Sentence-BERT) เป็นโมเดลที่พัฒนาขึ้นจาก BERT โดยเน้นไปที่การสร้างเวกเตอร์ที่แสดงความหมายของประโยค (sentence embeddings) มีการออกแบบมาให้รองรับการใช้งานที่ต้องการประสิทธิภาพสูง ซึ่งสามารถนำไปใช้ในงานที่เกี่ยวข้องกับการเปรียบเทียบข้อความ เช่น การหาความคล้ายคลึงระหว่างประโยค, การดึงข้อมูล (Information Retrieval), การจับคู่คำถาม-คำตอบ (Question-Answer Matching)

2.2. งานที่เกี่ยวข้อง

2.2.1. ระบบเทียบโอนรายวิชาและผลการเรียนด้วยการทำเหมืองข้อความ

ผู้จัดทำ: จรรยา แหยมเจริญ, ยงยุทธ ช่อจำปา, สุรพล คล้ายภมร, สายยาใจ พึ่งประชาชน

วัตถุประสงค์:

- เพื่อพัฒนาระบบเทียบโอนรายวิชาและผลการเรียนโดยประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อความ
- เพื่อพัฒนาเว็บแอปพลิเคชันสำหรับให้ผู้มีหน้าที่รับผิดชอบการเทียบโอนรายวิชาและคณะ/สาขาวิชาใช้งาน

วิธีการ:

การพัฒนาอัลกอริทึมการทำเหมืองข้อความ ประกอบด้วย

- อัลกอริทึมในการวิเคราะห์และคัดแยกคำศัพท์ของ คำอธิบายรายวิชาด้วยการทำ Word Tokenization และ Bag of Words ด้วยไลบรารี PythaiNLP

- อัลกอริทึมหาค่าความเหมือนของคำสำคัญ (Keyword) ของคำอธิบายรายวิชาที่ต้องการเทียบโอนมากับคำอธิบายรายวิชาของหลักสูตรวิทยาการคอมพิวเตอร์ด้วย Cosine Similarity พัฒนาด้วยภาษา PHP และพัฒนาเว็บแอปพลิเคชันด้วยภาษา Python และ PHP ด้วย Laravel Framework 5.4 บริหารจัดการข้อมูลด้วย MySQL หลักสูตรวิทยาการคอมพิวเตอร์ โดยเลือกกรณีศึกษาเป็นภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสยาม จากหนังสือเอกสารต่าง ๆ และแหล่งข้อมูลในอินเทอร์เน็ต

ข้อดี:

- ด้านการวิเคราะห์รายวิชาที่สามารถเทียบโอนได้ การตัดเพื่อหาคำสำคัญของคำอธิบายรายวิชาที่เป็นภาษาอังกฤษมีความคลาดเคลื่อนต่ำ ความถูกต้องแม่นยำสูง

ข้อเสีย:

- อัลกอริทึมสำหรับคัดแยกคำศัพท์เพื่อใช้เป็นคีย์เวิร์ดสำหรับเปรียบเทียบหาความเหมือนของคำอธิบายรายวิชาที่ต้องการเทียบโอนกับรายวิชาของหลักสูตรฯ โดยค่าความเหมือนที่ได้จากการทดสอบรายวิชาทางด้านการเขียนโปรแกรมคอมพิวเตอร์สูงสุดอยู่ระหว่าง 0.5 – 0.7 ซึ่งผลลัพธ์มีความคลาดเคลื่อนสูง บางรายวิชาไม่สามารถทำการเทียบโอนได้เนื่องจากเป็นรายวิชาที่ไม่มีความเกี่ยวข้องกัน
- คำอธิบายรายวิชาที่เป็นภาษาไทยมีความคลาดเคลื่อนสูงกว่าคำอธิบายรายวิชาที่เป็นภาษาอังกฤษ ซึ่งมีความถูกต้องแม่นยำน้อย

2.2.2. การพัฒนาระบบบริหารจัดการงานเทียบโอนรายวิชาเพื่อเพิ่มประสิทธิภาพการทำงานในสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏมหาสารคาม

ผู้จัดทำ: ปริญา ทองคำ

วัตถุประสงค์:

- เพื่อศึกษาปัญหาจากกระบวนการเทียบโอนรายวิชาในสาขาวิชาคอมพิวเตอร์ธุรกิจ
- เพื่อพัฒนาระบบสารสนเทศเพื่อจัดการรายวิชาเทียบโอน
- เพื่อศึกษาประสิทธิภาพของระบบสารสนเทศเพื่อการจัดการรายวิชาเทียบโอน

วิธีการ:

- ผู้วิจัยได้ดำเนินการสร้างและพัฒนาแบบสอบถามในการเก็บรวบรวมข้อมูลจากนักศึกษาสาขาวิชาคอมพิวเตอร์ธุรกิจ ผ่านแบบสอบถามเพื่อศึกษาสภาพปัญหาจริงของกระบวนการเทียบโอนรายวิชา ที่นักศึกษาจะต้องดำเนินการให้แล้วเสร็จภายในภาคเรียนสุดท้ายในหลักสูตร
- สร้างเครื่องมือที่จะใช้สำหรับเก็บข้อมูลจากนักศึกษา โดยมีการตรวจสอบคุณภาพเครื่องมือ และให้ผู้เชี่ยวชาญตรวจสอบความถูกต้องเพื่อเพิ่มความน่าเชื่อถือ
- การตรวจสอบคุณภาพเครื่องมือ

- นำแบบสอบถามที่สร้างขึ้นไปทดลองใช้กับนักศึกษาในสาขาวิชาคอมพิวเตอร์ธุรกิจ จำนวน 30 คน
 - การหาค่าอำนาจจำแนกของแบบสอบถามเป็นรายข้อ (Discriminant Power) ใช้เทคนิค Item-total Correlation
 - การหาค่าความเชื่อมั่น (Reliability Test) ของแบบสอบถามเป็นรายด้าน โดยใช้ค่าสัมประสิทธิ์แอลฟา (Alpha Coefficient) ตามวิธีของครอนบาค (Cronbach)
- จากนั้นทำการพัฒนาระบบสารสนเทศในการเทียบโอนรายวิชาตามขั้นตอนดังนี้
- ออกแบบฐานข้อมูลเพื่อใช้เก็บข้อมูล
 - ออกแบบฟอร์มรับข้อมูลและรายงานข้อมูลจากระบบ
 - เขียนโปรแกรมให้ระบบสามารถประมวลผลได้
 - ทดสอบและแก้ไขตามผลการประเมินจากแบบสอบถาม

ข้อดี:

- จากแบบสอบถามความคิดเห็นทำให้ทราบถึงปัญหาในกระบวนการเทียบโอนรายวิชาในสาขาวิชาคอมพิวเตอร์ธุรกิจในด้านต่าง ๆ จากข้อมูลเหล่านี้สามารถนำไปใช้เพื่อพัฒนาและปรับปรุงระบบการเทียบโอนรายวิชาให้ดีขึ้นได้

ข้อเสีย:

- งานวิจัยชิ้นนี้เป็นต้นแบบระบบสารสนเทศที่ใช้ภายใน แต่ถ้าหากต้องการนำไปใช้กับสาขาวิชาอื่น หรือคณะอื่น ๆ ที่มีความแตกต่างกันของหลักสูตร อาจต้องมีการปรับเปลี่ยนบางฟังก์ชันเพื่อให้ครอบคลุมทุกส่วนงาน เช่น ผู้ดูแลระบบจะต้องเพิ่มสาขาวิชา เพิ่มคณะลงในฐานข้อมูลได้ รวมทั้งสามารถกำหนดสิทธิ์ในการเข้าถึงข้อมูลจากหน้าจอระบบ เพื่อความสะดวกและเป็นมาตรฐาน

2.3. ตารางเปรียบเทียบโครงการกับงานอื่นที่มีอยู่ในปัจจุบัน

ตารางที่ 2-1 ตารางเปรียบเทียบโครงการกับงานอื่น

ลักษณะสำคัญ (Feature)	ระบบเทียบโอน รายวิชาและ ผลการเรียนด้วยการ ทำเหมืองข้อความ	การพัฒนาระบบบริหารจัดการ งานเทียบโอนรายวิชาเพื่อเพิ่ม ประสิทธิภาพการทำงานใน สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏ มหาสารคาม	โครงการของ ผู้จัดทำ
Data Analysis	/		/
Cosine Similarity	/		/
Word Tokenization	/		/
Bag of Words	/		
Contextual Embedding			/
Website Application	/	/	/
Actionable Insight		/	/
Network graph			/
Public Deploy			/

บทที่ 3 การดำเนินงานโครงการ

3.1. การรวบรวมข้อมูล (Data Collection/ Data Acquisition)

เป็นขั้นตอนการรวบรวมข้อมูล สำหรับนำไปวิเคราะห์

3.1.1. แหล่งข้อมูล (Data Sources)

ทำการดึงข้อมูลจากเว็บไซต์สำนักงานทะเบียนนักศึกษา มหาวิทยาลัยธรรมศาสตร์ ที่ใช้สำหรับค้นหารายวิชาต่าง ๆ (https://web.reg.tu.ac.th/registrar/class_info.asp?lang=en) โดยจะทำการเก็บรวบรวมข้อมูลจากคณะที่สนใจ ซึ่งมีทั้งหมด 5 คณะด้วยกัน ได้แก่

1. คณะวิศวกรรมศาสตร์
2. คณะวิทยาศาสตร์และเทคโนโลยี
3. วิทยาลัยสหวิทยาการ
4. วิทยาลัยนวัตกรรม
5. สถาบันเทคโนโลยีนานาชาติสิรินธร(SIIT)

โดยจะเก็บรายวิชาในภาคการศึกษาที่ 1/2566 และ 2/2566 ในระดับปริญญาตรี ทั้งโครงการปกติและโครงการพิเศษ ซึ่งเราจะเลือกรายละเอียดที่ต้องการผ่านหน้าเว็บไซต์แล้วกด search ซึ่งเราจะทำการเก็บข้อมูลของแต่ละคณะ โดยจะเก็บ 4 ส่วน ดังนี้

- รหัสรายวิชา (Course Code)
- คำอธิบายรายวิชา (Course Description)
- รายวิชาบังคับก่อน (Disciplinary Course)
- รายวิชาต่อเนื่อง (Consecutive Course)

3.1.2. การดึงข้อมูลจากเว็บไซต์ (Web Scraping)

เราใช้ Selenium ซึ่งเป็นเครื่องมือทำให้สามารถเขียนคำสั่งเพื่อทำงานบน browser ผ่านชุดคำสั่ง Python ได้ ทำให้สามารถออกแบบการทำงานต่าง ๆ กับเว็บไซต์ และสร้างเป็นระบบทำงานอย่างอัตโนมัติได้ เนื่องจากเว็บไซต์มีการสร้างเนื้อหาแบบเปลี่ยนแปลงได้ (Dynamic Websites) คือเว็บไซต์ที่ต้องให้ผู้ใช้งานมีพฤติกรรมร่วม เช่น ต้องเลื่อนเมาส์ รอกข้อมูลโหลด หรือกดปุ่มใด ๆ เพื่อให้โหลดข้อมูลที่ต้องการ ซึ่ง Selenium จะช่วยส่งการทำงานเหล่านี้ไปยังเว็บไซต์ได้อัตโนมัติ โดยขั้นตอนในการดึงข้อมูล มีดังนี้

1. ทำการเลือกข้อมูลของคณะ ภาควิชา การศึกษา ปีการศึกษา และระดับการศึกษาที่เราสนใจ แล้วคลิกที่ปุ่ม search

รูปที่ 3-1 เว็บไซต์สำนักงานทะเบียนนักศึกษา มหาวิทยาลัยธรรมศาสตร์

2. เมื่อคลิกปุ่ม search แล้วจะขึ้นรายวิชาทั้งหมดของคณะนั้น ๆ ซึ่งมีจำนวนค่อนข้างเยอะ จึงมีหลายหน้าเว็บ โดยจะมีปุ่ม next page เพื่อกดดูหน้าถัด ๆ ไป

Campus	Index	Course Code	Course Name	Unit of Credit	Section	Time	Classrooms	Seat Quota	remain	Status
RANGSET	2561	AM101	CALCULUS AND ITS APPLICATIONS 1	3	020001		101-10100 09:00-11:00 (Su)	70	22	Standard
RANGSET	2561	AM101	CALCULUS AND ITS APPLICATIONS 1	3	020002		101-10100 09:00-11:00 (Su)	9999	9913	Standard
RANGSET	2566	AM101	CALCULUS AND ITS APPLICATIONS 1	3	020001		101-10100 09:00-11:00 (Su)	70	22	Standard
RANGSET	2566	AM101	CALCULUS AND ITS APPLICATIONS 1	3	020002		101-10100 09:00-11:00 (Su)	9999	9913	Standard
RANGSET	2561	AM102	CALCULUS AND ITS APPLICATIONS 2	3	010001		101-10100 14:30-16:30 (Su)	9999	9966	Standard
RANGSET	2556	AM200	DISCRETE MATHEMATICS	3	020001		101-10100 09:00-11:00 (Su)	9999	9998	Standard
RANGSET	2561	AM200	DISCRETE MATHEMATICS	3	020001		101-10100 09:00-11:00 (Su)	9999	9943	Standard
RANGSET	2561	AM201	DIFFERENTIAL EQUATIONS FOR APPLIED MATHEMATICS	3	090001		101-10100 14:30-16:30 (Su)	9999	9964	Standard
RANGSET	2561	AM241	INTRODUCTION TO DATA SCIENCE AND ITS APPLICATIONS	3	070001		101-10100 12:00-14:00 (Su)	9999	9947	Standard
RANGSET	2561	AM260	APPLIED MATHEMATICS PRE-PROJECT	2	820001		101-10100 12:00-14:00 (Su)	50	8	Standard
RANGSET	2556	AM301	INTRODUCTION TO MODERN ALGEBRA	3	000001		101-10100 09:00-11:00 (Su)	9999	9998	Standard
RANGSET	2561	AM311	MATHEMATICAL ANALYSIS FOR APPLIED MATHEMATICS	3	020001		101-10100 09:00-11:00 (Su)	9999	9974	Standard
RANGSET	2561	AM320	INTEGRAL PROGRAMMING AND NETWORK MODELS	3	100001		101-10100 14:30-16:30 (Su)	9999	9971	Standard

รูปที่ 3-2 หน้าเว็บไซต์เมื่อทำการเลือกรายละเอียดต่าง ๆ

3. ทำการคลิกไปที่รหัสรายวิชาเพื่อดูรายละเอียดของวิชานั้น ๆ

AM102 CALCULUS AND ITS APPLICATIONS 2	
แคลคูลัสและการประยุกต์ 2	
Faculty	Faculty of Science and Technology
department	-
unit of credit	3
disciplinary course(s):	AM101
consecutive course(s):	AM201, AM311, AM327, MA316, MA351, MA412, ST226, ST321, ST327
course description	
Transformation of coordinate systems, polar coordinates and graphing, vectors, lines and planes in 3D- space, functions of several variables and surface plot, limits and continuity of several variables functions, partial derivatives, higher-order partial derivatives, implicit functions theorem, applications of partial derivatives in optimization, multiple integrations, differential vector calculus, introduction to line integrals and surface integrals, Gauss's Theorem, Green's Theorem and Stokes' Theorem, software programme for calculus of several variable functions.	

รูปที่ 3-3 ตัวอย่างหน้าเว็บข้อมูลรายวิชา

4. ซึ่งเราจะใช้ Selenium ทำงานตั้งแต่ขั้นตอนที่ 2 เป็นต้นไป โดยเราจะกำหนด element ส่วนเนื้อหาที่เราต้องการ ซึ่งก็คือ course code, disciplinary course(s), consecutive course(s) และ course description เพื่อให้ Selenium ไปดึงข้อมูลในส่วนนั้น ๆ แล้วให้ Selenium ทำงานวนลูปไปเรื่อย ๆ และเมื่อเจอปุ่ม Next page ให้คลิกแล้วไปที่หน้าถัดไป และดึงข้อมูลเหมือนเดิม

3.2. การเตรียมข้อมูลเบื้องต้น (Data Preprocessing / Data Cleaning)

เป็นขั้นตอนที่สำคัญในการเตรียมข้อมูลให้เหมาะสมและมีคุณภาพก่อนที่จะนำไปวิเคราะห์และสร้างโมเดล เมื่อเราดึงข้อมูลโดยใช้ Selenium เรียบร้อยแล้ว เราจะนำข้อมูลที่ได้มาสร้าง DataFrame และส่งข้อมูลผ่าน API เพื่อสร้าง Database แล้วนำมาดำเนินการดังต่อไปนี้

```
dict = {"code": list_codes, "desc": list_desc, "add1": add1, "add2": add2, "add3": add3}
dataframe = pd.DataFrame(dict)
dataframe.head()
```

	code	desc	add1	add2	add3
0	AM101	Limit and continuity of single variable functi...	[disciplinary course(s): ไม่นับหน่วยกิต ให้อีก...	[consecutive course(s): AM102, ST326]	[]
1	AM101	Limit and continuity of single variable functi...	[disciplinary course(s): ไม่นับหน่วยกิต ให้อีก...	[consecutive course(s): AM102, ST326]	[]
2	AM101	Limit and continuity of single variable functi...	[disciplinary course(s): ไม่นับหน่วยกิต ให้อีก...	[consecutive course(s): AM102, ST326]	[]
3	AM101	Limits and continuity of single variable funct...	[disciplinary course(s): หมายเหตุ : ไม่นับหน่วยกิต...	[consecutive course(s): AM102, AM201, ST326, S...	[]
4	AM101	Limits and continuity of single variable funct...	[disciplinary course(s): หมายเหตุ : ไม่นับหน่วยกิต...	[consecutive course(s): AM102, AM201, ST326, S...	[]

รูปที่ 3-4 Scraping Data DataFrame

3.2.1. Convert Data Type

ทำการแปลงประเภทของข้อมูลในคอลัมน์ทั้งหมดของชุดข้อมูลของรายวิชาทั้ง 5 คณะ โดยกำหนดให้ทุก Features add1, add2 และ add3 คือ disciplinary course(s), consecutive course(s) และค่า NaN ตามลำดับ จากนั้นทำการแปลงประเภทข้อมูลให้เป็น string

```
dataframe[["add1", "add2", "add3"]] = dataframe[["add1", "add2", "add3"]].astype(str)
```

รูปที่ 3-5 Convert Data Type

3.2.2. เก็บชุดข้อมูลใน Google Sheet

ทำการเก็บชุดข้อมูลของรายวิชาแต่ละคณะ โดยเลือกกระตักการศึกษาปริญญาตรี ภาคการศึกษาที่ 1/2566 และ 2/2566 รวมถึงเลือกโครงการทั้งโครงการปกติและโครงการพิเศษ ซึ่งจะได้ชุดข้อมูลแต่ละภาคการศึกษา โครงการ และคณะ โดยทางคณะผู้จัดทำได้เพิ่ม column “ma” ซึ่งเป็น column ที่บอกคณะ เพื่อใช้ในขั้นตอนการสร้างโมเดล

	A	B	C	D	E	F	G
1	code	desc	add1	add2	add3	ma	
2	AM101	Limit and contin	[disciplinary course(s) ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษาหรือสอบได้ ค. 111 หรือ ค. 211 หรือ ค. 216 หรือ ค. 218]	[consecutive coi []		sci	
3	AM101	Limit and contin	[disciplinary course(s) ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษาหรือสอบได้ ค. 111 หรือ ค. 211 หรือ ค. 216 หรือ ค. 218]	[consecutive coi []		sci	
4	AM101	Limit and contin	[disciplinary course(s) ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษาหรือสอบได้ ค. 111 หรือ ค. 211 หรือ ค. 216 หรือ ค. 218]	[consecutive coi []		sci	
5	AM101	Limits and contr	[disciplinary course(s) หมายถึง : ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษาหรือสอบได้ ค. 111 หรือ ค. 113 หรือ ค. 211 หรือ ค. 216 หรือ ค. 21	[consecutive coi []		sci	
6	AM101	Limits and contr	[disciplinary course(s) หมายถึง : ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษาหรือสอบได้ ค. 111 หรือ ค. 113 หรือ ค. 211 หรือ ค. 216 หรือ ค. 21	[consecutive coi []		sci	
7	AM102	Transformation c	[disciplinary course(s) AM101]	[consecutive coi []		sci	
8	AM200	Logic, algebra of	[disciplinary course(s) หมายถึง 1. วิชาที่เทียบเท่า (EQUIVALENT) กับวิชา คผ. 101 โครงสร้างแบบไม่ต่อเนื่อง 2. ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษา	[consecutive coi []		sci	
9	AM200	Logic, algebra of	[disciplinary course(s) 1. วิชาที่เทียบเท่า (equivalent) กับวิชา คผ. 101 โครงสร้างแบบไม่ต่อเนื่อง 2. ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษา	[consecutive coi []		sci	
10	AM201	First order differ	[disciplinary course(s) AM102 or inParallel degree/ Dual degree AM102]	[consecutive coi []		sci	
11	AM241	Computer progr	[consecutive course(s) ST277]	[]	[]	sci	
12	AM260	Software packag	[disciplinary course(s) ศึกษาชั้นปีที่ 2 ขึ้นไป]	[consecutive coi []		sci	
13	AM301	Sets, mapping a []		[]	[]	sci	
14	AM311	Real number sys	[disciplinary course(s) AM102 and MA221]	[consecutive coi []		sci	
15	AM326	Integer program	[disciplinary course(s) AM321]	[]	[]	sci	
16	AM328	Market mechan	[disciplinary course(s) CS103]	[]	[]	sci	
17	AM341	Data mining met	[disciplinary course(s) MA332 and ST211 or inMA131 and ST211 or inMA332 and ST216 or inMA131 and ST216]	[]	[]	sci	
18	AM346	Finite state mac	[disciplinary course(s) AM200 or inCS101]	[]	[]	sci	
19	AM347	Mathematical-pr	[disciplinary course(s) AM246 or inCS213]	[]	[]	sci	
20	AM411	Meaning and val	[disciplinary course(s) AM200 and MA332 and ST226 หรือได้รับอนุมัติจากผู้บริหาร]	[]	[]	sci	
21	AM428	Introduction to p	[disciplinary course(s) AM321 or Approved by department/instructor]	[]	[]	sci	
22	AM429	Roles of supply	[disciplinary course(s) AM321]	[]	[]	sci	
23	AM446	Graphic devices	[disciplinary course(s) CS103 and MA332 or inCS103 and MA131]	[]	[]	sci	

รูปที่ 3-6 ชุดข้อมูลใน Google Sheet

3.2.3. String Processing

ทุก Features ได้ทำการแก้ไขเป็นตัวพิมพ์เล็ก (lowercase) เพื่อแก้ปัญหา Case Sensitive

```
for i in dataframe.columns:
    dataframe[f"{i}"] = dataframe[f"{i}"].str.strip().str.lower()
```

รูปที่ 3-7 lowercase

และเนื่องจากคำอธิบายรายวิชา (Course Description) มีบางรายการซ้ำกัน รวมถึงมี . (full stop), , (comma) และ ช่องว่าง(Space) โดยใช้ Strip ในการตัดออก

```
dataframe["desc"] = dataframe["desc"].str.replace(r"[.,]", "", regex=True).str.strip()
```

รูปที่ 3-8 Course Description Standard

3.2.4. Drop NaN and Duplicate

เนื่องจากรหัสรายวิชา (Course code) มีค่าซ้ำกันเนื่องจากอาจารย์ผู้สอนคนละท่านและต่างหลักสูตรหรือ section รวมถึงรหัสรายวิชา (Course code) และคำอธิบายรายวิชา (Course Description) ที่มีค่า NaN เราจึงทำการ drop ทิ้ง

```
dataframe.drop_duplicates(subset=["code"],inplace=True)
dataframe.dropna(subset=["description","code"],inplace=True)
```

รูปที่ 3-9 Drop NaN and Duplicate

3.2.5. Course Code Standard

แม้จะเลือกใช้เว็บไซต์ในการดึงข้อมูลเป็นภาษาอังกฤษแล้ว แต่ยังพบว่ายังมีบางส่วนที่รหัสรายวิชาเป็นภาษาไทยในส่วนของ disciplinary course(s) และ consecutive course(s) จึงทำการเปลี่ยนเป็นภาษาอังกฤษทั้งหมด โดยเปลี่ยนดังนี้

- ค (คณิต): MA
- คพ (วิทยาการคอมพิวเตอร์): CS
- ส (สถิติ): ST
- ทก (เทคโนโลยีการเกษตร): AT
- คป (คณิตศาสตร์ประยุกต์): AM
- ทช (เทคโนโลยีชีวภาพ): BT
- วท (วิทยาศาสตร์): SC
- คม (คอมพิวเตอร์): CM

```
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("คพ","cs")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("ทก","at")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("คป","am")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("ทช","bt")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("วท","sc")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("คม","cm")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("ค","ma")
dataframe["disciplinary course(s)"] = dataframe["disciplinary course(s)"].str.replace("ส","st")

dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("คพ","cs")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("ทก","at")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("คป","am")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("ทช","bt")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("วท","sc")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("คม","cm")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("ค","ma")
dataframe["consecutive course(s)"] = dataframe["consecutive course(s)"].str.replace("ส","st")
```

รูปที่ 3-10 Course Code Standard

3.2.6. RegEx

ใช้ RegEx เพื่อเลือกเพียงรายวิชาบังคับก่อน (disciplinary course) และ รายวิชาต่อเนื่อง (consecutive course) จาก column “disciplinary course(s)” และ “consecutive course(s)” ที่เป็นประโยคข้อความ โดยดึงเพียง course code จากทั้งสอง column มาเพิ่มเป็น column ใหม่ที่มีชื่อว่า ‘cancel_course’ เนื่องจากรายวิชาบังคับก่อนและรายวิชาต่อเนื่องอาจมีความคล้ายคลึง แต่ไม่สามารถเทียบโอนได้จึงต้องเพิ่ม column นี้ รวมถึงเปลี่ยนชื่อ column อื่น ๆ ให้เหมาะสมกับเนื้อหาใน column นั้น ๆ

```
def extract_courses(row):
    pattern = r"[A-Za-z0-9]{1,2}\.? \d{3}"
    listt = []
    if isinstance(row, str):
        matches = re.findall(pattern, row)
        listt.extend(matches)
        return listt
    return None

dataframe["cancel_course"] = dataframe["disciplinary course(s)"].apply(extract_courses) + dataframe["consecutive course(s)"]
```

รูปที่ 3-11 RegEx

3.2.7. จัดเก็บชุดข้อมูลในรูปไฟล์ .csv

หลังจาก Data Cleaning เสร็จ เราได้ทำการจัดเก็บชุดข้อมูลในรูปแบบไฟล์ .csv เพื่อให้สะดวกในการใช้ในขั้นตอนการสร้างแบบจำลอง (Model Building) โดยมี column ดังต่อไปนี้ code (course code), description (course description), disciplinary course(s), consecutive course(s), faculty และ cancel_course

	code	description	disciplinary course(s)	consecutive course(s)	faculty	cancel_course
0	am100	basics concepts in applied mathematics mathematical modeling operations research and computational mathematics	[]	[]	sci	[]
1	am101	limit and continuity of single variable functions the derivative and its applications antiderivatives and indefinite integrals techniques of integrations definite integrals and applications of the definite integrals in geometry and physics improper integrals and convergence tests infinite series and convergence tests power series and theorem of taylor series software programme for calculus of single variable functions	['disciplinary course(s): ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษาหรือสอบได้ ma.111 หรือ ma.211 หรือ ma.216 หรือ ma.218']	['consecutive course(s): am102, st326']	sci	['ma.111', 'ma.211', 'ma.216', 'ma.218', 'am102', 'st326']
2	am102	transformation of coordinate systems polar coordinates and graphing vectors lines and planes in 3d- space functions of several variables and surface plot limits and continuity of several variables functions partial derivatives higher-order partial derivatives implicit functions theorem applications of partial derivatives in optimization multiple integrations differential vector calculus introduction to line integrals and surface integrals gauss's theorem green's theorem and stokes' theorem software programme for calculus of several variable functions	['disciplinary course(s): am101']	['consecutive course(s): am201, am311, am327, am316, ma351, ma412, st226, st321, st327']	sci	['am101', 'am201', 'am311', 'am327', 'ma316', 'ma351', 'ma412', 'st226', 'st321', 'st327']
3	am151	symbolic logic methods of proof proof of quantifier sentences real number system properties of real numbers sets properties and theorems of sets relations equivalence relations partial order relations functions and types of functions	['consecutive course(s): am211, am457, ma332, ma337, ma338, ma366, ma368']	[]	sci	['am211', 'am457', 'ma332', 'ma337', 'ma338', 'ma366', 'ma368']
4	am200	logic algebra of sets relations and functions recurrence relations generating functions basic concepts in graph theory number representation in computer boolean algebra and combinatorial circuits finite-state machines finite automata basic concepts in algebraic system note : 1 this subject is equivalent to cs 101 discrete structures 2 there is no credit for students who currently taking or have earned credits of cs 101	['disciplinary course(s): 1.วิชาที่เทียบเท่า(equivalent)กับวิชา cs.101 โทมัสสร้างแบบไม่ต่อเนื่อง 2.ไม่นับหน่วยกิตให้ผู้ที่กำลังศึกษา หรือสอบได้cs.101']	['consecutive course(s): am346, am411, am448, am451, ma338, ma366, ma367']	sci	['cs.101', 'cs.101', 'am346', 'am411', 'am448', 'am451', 'ma338', 'ma366', 'ma367']

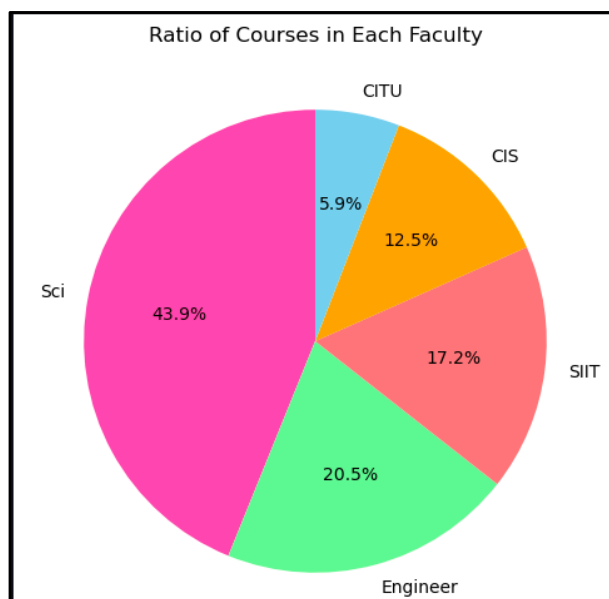
รูปที่ 3-12 DataFrame

3.3. การวิเคราะห์ภาพรวมข้อมูล (Exploratory Data Analysis / Data Visualization)

เป็นขั้นตอนในการวิเคราะห์ข้อมูล เพื่อแสดงภาพรวมและทำให้เข้าใจข้อมูลมากขึ้น โดยได้ดำเนินการดังนี้

3.3.1. Distribution

ตรวจสอบสัดส่วนของข้อมูลจำนวนรายวิชาด้วยการ plot กราฟแท่ง โดยแยกเป็นแต่ละคณะ ดังนี้ คณะวิทยาศาสตร์และเทคโนโลยี (สีชมพู), คณะวิศวกรรมศาสตร์ (สีเขียว), สถาบันเทคโนโลยี นานาชาตีสิรินคร(SIIT) (สีแดง), วิทยาลัยสหวิทยาการ (สีส้ม) และวิทยาลัยนวัตกรรม (สีฟ้า) โดยจำนวนรายวิชาของแต่ละคณะคิดเป็นเปอร์เซ็นต์ได้ดังนี้

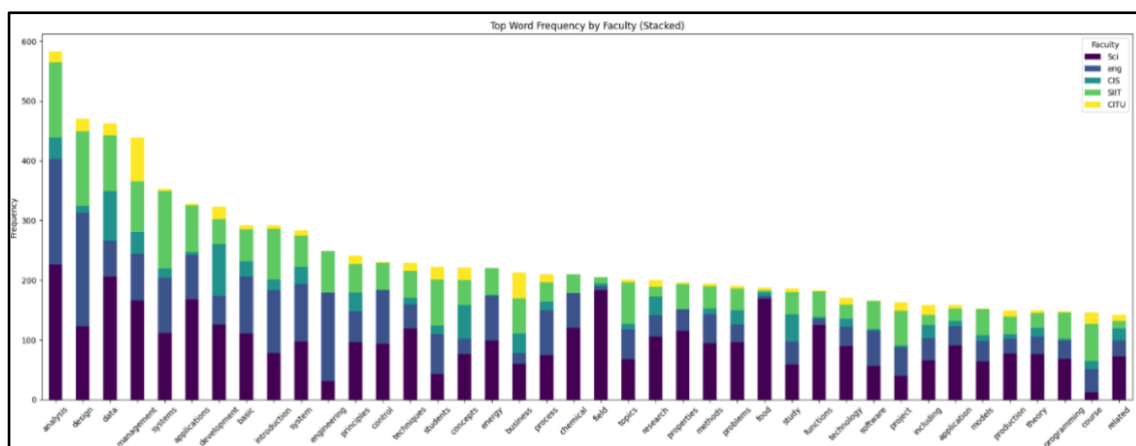


รูปที่ 3-13 pie chart แสดงจำนวนรายวิชาทั้งหมดของ 5 คณะ (คิดเป็น %)

จากpie chart ข้อมูลของจำนวนรายวิชาในแต่ละคณะ จะเห็นได้ว่ามีสัดส่วนของรายวิชาในคณะวิทยาศาสตร์มากที่สุด เนื่องจากมีสาขาวิชามากที่สุดจึงทำให้สัดส่วนมากที่สุดเมื่อเทียบกับคณะอื่น ๆ

3.3.2. Top Word Frequency by Faculty (Stacked bar chart)

การหาความถี่ของคำที่พบมากที่สุดในแต่ละคณะ โดยใช้กราฟแท่งแบบต่อ (Stacked bar chart) ในการแสดงผล



รูปที่ 3-14 กราฟแท่งแบบต่อแสดงค่าความถี่ของคำที่พบมากที่สุดแต่ละคณะ

จากกราฟพบว่าคำที่พบมากที่สุดคือ “analysis” โดยพบมากในคณะวิทยาศาสตร์และเทคโนโลยี และคณะวิศวกรรมศาสตร์ในสัดส่วนใกล้เคียงกัน สัดส่วนของคำที่พบมากที่สุดรองลงมาและใกล้เคียงกันคือคำว่า ‘design’ และ ‘data’ ซึ่งคำว่า ‘design’ พบมากในวิทยาลัยสหวิทยาการ ส่วนคำว่า ‘data’ พบมากในคณะวิทยาศาสตร์และเทคโนโลยี

3.4. การสร้างแบบจำลอง (Model Building)

3.4.1. Embeddings

หลังจากการทำการเตรียมข้อมูลเบื้องต้น (Data Preprocessing / Data Cleaning) และได้ทำการจัดเตรียมชุดข้อมูล (dataset) เพื่อใช้ในขั้นตอนนี้ เราได้ทำการสร้างฟังก์ชันเพื่อสร้าง Embeddings (create_embeddings1, create_embeddings2 และ create_embeddings3) แล้วดึงข้อมูลจากคอลัมน์ description ของ DataFrame มาใช้ โดยเราได้เลือกใช้โมเดลที่เป็น SentenceTransformer ในการสร้างเวกเตอร์ขนาดคงที่สำหรับทั้งประโยค ช่วยลดความซับซ้อนในการสร้างและนำไปใช้ต่อ โดยเวกเตอร์นี้จะสรุปความหมายของประโยคทั้งหมดเอาไว้เป็น Embedding ระดับประโยค (Sentence-Level Embedding) จับความหมายเชิง semantic ของประโยคได้อย่างมีประสิทธิภาพ แม้คำจะต่างกันแต่ความหมายคล้ายกัน และยังมีข้อดีในด้านความเร็วและประสิทธิภาพ โดยเราเลือกใช้โมเดล 3 ตัว คือ all-MiniLM-L6-v2 ถูกปรับแต่งมาเพื่อการใช้งานที่ต้องการความเร็วเป็นหลัก, paraphrase-mpnet-base-v2 ถูก fine-tune มาเพื่อจับความคล้ายคลึงของข้อความ (Paraphrase tasks) โดยเฉพาะ และ all-mpnet-base-v2 ปรับแต่งมาเพื่อการทำงานทั่วไป (General Purpose Sentence Embedding) Generalized Model ไม่ได้เน้นเฉพาะ Paraphrase Tasks แต่ให้ embedding ที่มีความหมายดี โดยฟังก์ชันจะ return ผลลัพธ์ embeddings ที่ได้กลับมา แล้วนำผลลัพธ์มาบันทึกใน DataFrame

```
model1 = SentenceTransformer("sentence-transformers/all-MiniLM-L6-v2")
model2 = SentenceTransformer('paraphrase-mpnet-base-v2')
model3 = SentenceTransformer("all-mpnet-base-v2")

def [7]:
def create_embeddings1():
    dis = list(dataframe["description"])
    embeddings = model1.encode(dis)
    return embeddings
dataframe["embed1"] = create_embeddings1().tolist()

def create_embeddings2():
    dis = list(dataframe["description"])
    embeddings = model2.encode(dis)
    return embeddings
dataframe["embed2"] = create_embeddings2().tolist()

def create_embeddings3():
    dis = list(dataframe["description"])
    embeddings = model3.encode(dis)
    return embeddings
dataframe["embed3"] = create_embeddings3().tolist()

embeddings1 = np.array(dataframe["embed1"].tolist())
embeddings2 = np.array(dataframe["embed2"].tolist())
embeddings3 = np.array(dataframe["embed3"].tolist())
```

รูปที่ 3-15 SBERT Model

embed1	embed2	embed3
[-0.03120432049036026,	[-0.10621019452810287,	[-0.05130860581994057,
-0.023463578894734383,	0.07966027408838272,	-0.060512710362672806,
-0.059237826615571976,	-0.0558156780898571,	-0.025702929124236107,
-0.04906650632619858,	0.03535414859652519,	-0.017236024141311646,
-0.14708583056926727,	-0.03739959001541138,	-0.038748372346162796,
-0.12325919419527054,	-0.098751001060009,	-0.017528556287288666,
0.028603173792362213,	0.025458503514528275,	0.013874592259526253,
0.0572102852165699,	-0.05442697927355766,	-0.017771882936358452,
-0.002235500840470195,	-0.28438296914100647,	-0.1012776717543602,
0.01441214233636856,	0.08761554956436157,	0.010542075149714947,
-0.07094401121139526,	0.10207586735486984,	0.007238856051117182,
0.06099458411335945,	0.09625472128391266,	-0.024972153827548027,
0.05436413735151291,	0.03691395744681358,	0.040836773812770844,
0.015640532597899437,	0.037167586386203766,	0.05128929764032364,
0.018084801733493805,	-0.06262032687664032,	0.004058344755321741,
-0.00695278262719512,	-0.14318789541721344,	-0.06219536066055298,
-0.02605459839105606,	0.008872530423104763,	0.02055571973323822,
0.03367605060338974,	-0.038490742444992065,	-0.0068100616335868835,
-0.06808045506477356,	-0.19456075131893158,	-0.020049020648002625,
-0.06047604978084564,	0.09074630588293076,	-0.024581698700785637,
0.04407086223363876,	0.022933287546038628,	0.01401970349252224,
-0.06672877073287964,	0.15999357402324677,	-0.042191047221422195,
-0.08734912425279617,	0.13747066259384155,	-0.012145196087658405,
-0.00458286190405488,	0.14363226294517517,	0.02643175609409809,
0.014137336052954197,	0.03455844148993492,	-0.018343592062592506,
-0.02638586238026619,	-0.15276895463466644,	-0.032709378749132156,
0.004036597907543182,	-0.09958638995885849,	0.025912009179592133,
-0.011059576645493507,	-0.053255319595336914,	0.023654893040657043,
0.08288545161485672,	0.07561995834112167,	-0.03545462340116501,
-0.009969662874937057,	0.11504115909337997,	-0.01864059828221798,
-0.0857098251581192,	-0.03640494495630264,	0.03044602833688259,
0.010402277112007141,	-0.000673578295391053,	0.00481991795822978,
0.09473756700754166,	0.0295135248452425,	0.02876494638621807,
0.020847750827670097,	0.03212375193834305,	0.01749230921268463,
-0.0020401556976139545,	0.0031077335588634014,	1.4249295645640814e-06,
0.03116466850042343,	-0.012611215002834797,	-0.021548354998230934,
0.015065344050526619,	-0.0038778551388531923,	

รูปที่ 3-16 Embed Column

3.4.2. คำนวณและแสดง Cosine Similarity Network ระหว่างรายวิชาที่มี embeddings โดยมีจุดประสงค์เพื่อหาคู่รายวิชาที่มีความคล้ายคลึงกันภายใต้เงื่อนไขที่กำหนด

โดยเริ่มจาก คำนวณค่า Cosine Similarity ใน embed column สร้างเงื่อนไขเพื่อจับคู่ที่มีความคล้ายคลึงกัน เช่น faculty ต่างกัน เก็บ course code ที่มี description คล้ายคลึงกัน แล้วเพิ่มข้อมูลลงใน DataFrame โดยใช้ networkx สร้างกราฟโดยมีเงื่อนไขความคล้ายคลึงกัน (Edge) และทำทั้งหมดทุก model หลังจากนั้นบันทึกข้อมูลเป็นไฟล์ data_pairs.csv

```

def compute_similarity_matrix1(dataframe, threshold=0.85):
    embeddings = np.vstack(dataframe["embed1"].dropna().to_numpy())
    similarity_matrix = cosine_similarity(embeddings)

    pairs = []
    code_pairs = []
    for i in range(len(similarity_matrix)):
        for j in range(i + 1, len(similarity_matrix)):
            if similarity_matrix[i, j] > threshold:
                course_i = dataframe.iloc[i]["code"]
                course_j = dataframe.iloc[j]["code"]
                cancel_i = dataframe.iloc[i]["cancel_course"]
                cancel_j = dataframe.iloc[j]["cancel_course"]
                ii = dataframe.iloc[i]["faculty"]
                jj = dataframe.iloc[j]["faculty"]

                # and jj not in ii and ii not in jj
                if course_j not in cancel_i and course_i not in cancel_j and jj not in ii and ii not in jj:
                    pairs.append((i, j))
                    code_pairs.append((course_i, course_j))

    dataframe["valid_pairs1"] = dataframe["code"].apply(
        lambda x: [s for s in code_pairs if x in s]
    )
    dataframe["valid_pairs1_85"] = dataframe.apply(
        lambda row: [pair[0] if pair[1] == row["code"] else pair[1] for pair in row["valid_pairs1"] if row["code"] in pair],
        axis=1
    )

    codes = dataframe["code"].tolist()
    G = nx.Graph()
    G.add_edges_from([(codes[i], codes[j]) for i, j in pairs])

    plt.figure(figsize=(20, 15))
    pos = nx.spring_layout(G, seed=42)
    nx.draw_networkx_nodes(G, pos, node_size=300, node_color='lightblue')
    nx.draw_networkx_edges(G, pos, alpha=0.5, edge_color='gray')
    nx.draw_networkx_labels(G, pos, font_size=10, font_color="black")

    plt.title(f"Similarity Network ")
    plt.show()
    return code_pairs

dataframe["valid_pairs1_85"] = [[] for _ in range(len(dataframe))]
code_pairs1 = compute_similarity_matrix1(dataframe)

```

รูปที่ 3-17 Cosine Similarity

3.5. การนำแบบจำลองไปใช้งาน (Model Deployment)

ทางคณะผู้จัดทำได้พัฒนาเว็บไซต์เพื่ออำนวยความสะดวกในการดูแลผลลัพธ์ อีกทั้งยังสามารถใช้เป็นแนวทางในการเพิ่มวิชาที่สามารถเทียบโอนได้ระหว่างคณะภายในมหาวิทยาลัยธรรมศาสตร์ในอนาคต โดยมุ่งเน้นให้การใช้งานข้อมูลเป็นไปอย่างสะดวกและมีประสิทธิภาพสำหรับมหาวิทยาลัยที่ต้องการดูรายวิชาที่สามารถเทียบโอนได้ และนักศึกษาที่ต้องการย้ายคณะ ซึ่งในโครงการนี้จะครอบคลุมคณะจำนวน 5 คณะ ได้แก่ คณะวิศวกรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, วิทยาลัยสหวิทยาการ, วิทยาลัยนวัตกรรมการและสถาบันเทคโนโลยีนานาชาติสิรินธร(SIIT) โดยทางคณะผู้จัดทำได้ใช้ Streamlit framework ในการพัฒนา Web Application เพื่อให้ผู้ใช้งานสามารถเข้าถึงข้อมูลและทำการตรวจสอบการเทียบโอนรายวิชาได้อย่างมีประสิทธิภาพ ผ่านขั้นตอนการพัฒนาดังต่อไปนี้

3.5.1. การดึงข้อมูลมาใช้งาน

ข้อมูลที่ใช้ในการพัฒนาเว็บไซต์ได้มาจากไฟล์ data_pairs.csv ซึ่งเป็นฐานข้อมูลหลักที่รวบรวมรายละเอียดของรายวิชาที่มีโอกาสเทียบโอนกันได้ ระหว่างคณะต่าง ๆ ภายในมหาวิทยาลัย รวมถึงคำอธิบายรายวิชา (Course Description) ของแต่ละรายวิชา

3.5.2. การตั้งค่า Page Config และ Session State

เพื่อเพิ่มประสิทธิภาพในการแสดงผล เว็บไซต์ได้ถูกตั้งค่า Page Config เพื่อปรับแต่งการแสดงผลให้เหมาะสมกับผู้ใช้งาน พร้อมทั้งใช้ Session State ในการจัดการสถานะของข้อมูลภายในเว็บแอปพลิเคชัน ช่วยให้ระบบสามารถตอบสนองต่อการเปลี่ยนแปลงที่เกิดขึ้นจากการใช้งานได้อย่างราบรื่น

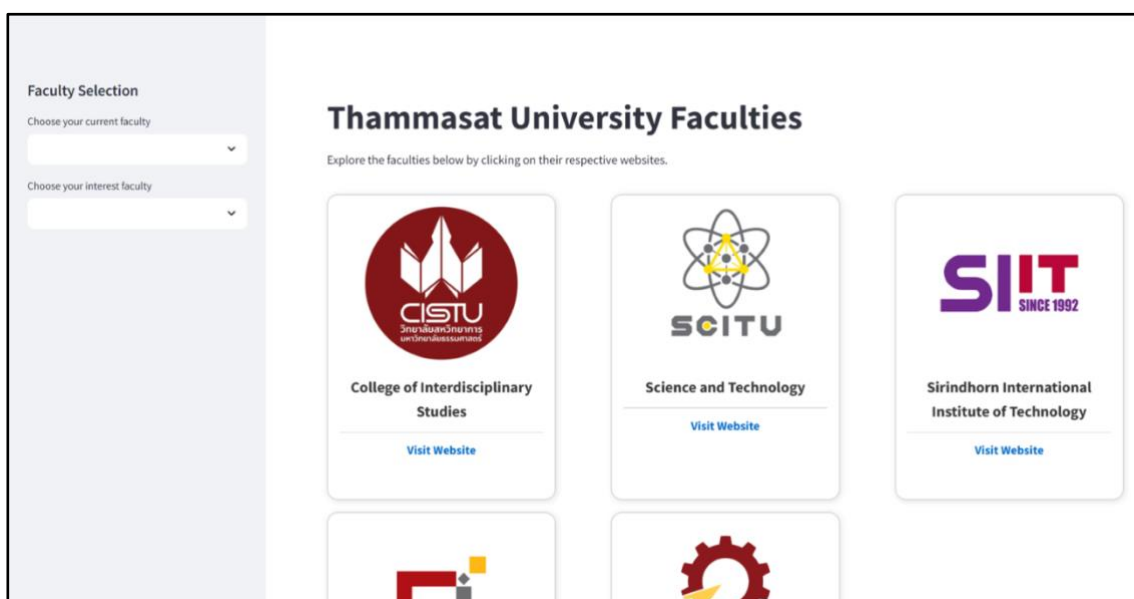
3.5.3. การออกแบบหน้า Home Page

หน้า Home Page ถูกออกแบบให้เป็นจุดเริ่มต้นของการใช้งาน โดยจะแสดงรายการคณะที่เราได้ทำการเก็บข้อมูลมา พร้อมทั้งลิงก์ไปยังเว็บไซต์อย่างเป็นทางการของคณะนั้น ๆ เพื่อให้ผู้ใช้งานสามารถศึกษาข้อมูลเพิ่มเติมได้อย่างสะดวก

นอกจากนี้ยังมีการ Sidebar เพื่อให้ผู้ใช้งานสามารถเลือกคณะปัจจุบัน และคณะที่ต้องการเทียบโอน ได้ง่ายขึ้น ช่วยลดความซับซ้อนในการค้นหาและการเปรียบเทียบข้อมูล

3.5.4. การตรวจสอบและแสดงผลรายวิชาที่เทียบโอนได้

เมื่อผู้ใช้งานเลือกคณะปัจจุบันและคณะที่ต้องการเทียบโอน ระบบจะดึงข้อมูลจากฐานข้อมูล และแสดงรายการรายวิชาที่สามารถเทียบโอนกันได้ พร้อมทั้งแสดงคำอธิบายรายวิชา (Course Description) ของรายวิชาทั้งสองฝั่งอย่างละเอียด ซึ่งจะช่วยให้ผู้ใช้งานสามารถตรวจสอบความสอดคล้องของรายวิชาได้อย่างชัดเจน



รูปที่ 3-18 Home Page

รูปที่ 3-19 Pair Course Page

โดยสรุปแล้ว เว็บไซต์นี้จะช่วยอำนวยความสะดวกให้มหาวิทยาลัยในการเพิ่มรายวิชาที่สามารถเทียบโอนกันได้ รวมถึงนักศึกษาที่ต้องการจะย้ายคณะ โดยจะสามารถตรวจสอบรายวิชาที่สามารถเทียบโอนได้ เพื่อใช้เป็นแนวทางในการตัดสินใจ โดยสามารถตรวจสอบข้อมูลการเทียบโอนรายวิชาได้อย่างรวดเร็วและถูกต้อง อีกทั้งยังช่วยลดความยุ่งยากในการค้นหาข้อมูล ด้วยการนำเทคโนโลยี Streamlit framework มาพัฒนาให้เป็นเครื่องมือที่มีความทันสมัย ใช้งานง่าย และตอบสนองต่อความต้องการของผู้ใช้งานได้อย่างมีประสิทธิภาพ อีกทั้งยังเป็นแนวทางในการสร้างธนาคารหน่วยกิตของมหาวิทยาลัยธรรมศาสตร์ในอนาคต

บทที่ 4 ผลการดำเนินงานโครงการ

4.1. Dataframe

ผลลัพธ์ที่ได้จากการวิเคราะห์และจับคู่รายวิชาที่สามารถเทียบโอนหน่วยกิตได้ เกิดขึ้นจากการคำนวณความคล้ายคลึงกันของคำอธิบายรายวิชา (Course Description) โดยใช้เทคนิค Cosine Similarity ในการวัดค่าความคล้ายคลึงของข้อความ และทำการคัดเลือกคู่รายวิชาที่มีค่า Cosine Similarity สูงกว่าเกณฑ์ที่กำหนด (Threshold) รวมถึงต้องอยู่ภายใต้เงื่อนไขที่กำหนดไว้ เช่น รายวิชาที่จับคู่กันจะต้องไม่เป็นวิชาที่อยู่ในคณะเดียวกัน เพื่อให้เกิดประสิทธิภาพสูงสุดในการเทียบโอนรายวิชาระหว่างคณะ

ผลลัพธ์ที่ได้จะถูกจัดเก็บในรูปแบบของ DataFrame ซึ่งเป็นโครงสร้างข้อมูลที่มีความยืดหยุ่นและเหมาะสมต่อการนำไปใช้งานต่อในอนาคต โดยภายใน DataFrame จะประกอบไปด้วย คู่รายวิชาที่สามารถเทียบโอนกันได้ พร้อมทั้งรายละเอียดที่สำคัญคือ รหัสวิชา (code), คำอธิบายรายวิชา (description), คณะ (faculty), รายวิชาบังคับก่อนและรายวิชาต่อเนื่อง (cancel_course) และคู่รายวิชาที่สามารถเทียบโอนได้ (valid_pairs)

	code	description	faculty	cancel_course	valid_pairs1_65
0	am100	basics concepts in applied mathematics mathematical modeling operations research and computational mathematics	sci	[]	[ies321]
1	am101	limit and continuity of single variable functions the derivative and its applications antiderivatives and indefinite integrals techniques of integrations definite integrals and applications of the definite integrals in geometry and physics improper integrals and convergence tests infinite series and convergence tests power series and theorem of taylor series software programme for calculus of single variable functions	sci	['ma.111', 'ma.211', 'ma.216', 'ma.218', 'am102', 'st326']	[mas116]
2	am102	transformation of coordinate systems polar coordinates and graphing vectors lines and planes in 3d- space functions of several variables and surface plot limits and continuity of several variables functions partial derivatives higher-order partial derivatives implicit functions theorem applications of partial derivatives in optimization multiple integrations differential vector calculus introduction to line integrals and surface integrals gauss's theorem green's theorem and stokes' theorem software programme for calculus of several variable functions	sci	['am101', 'am201', 'am311', 'am327', 'ma316', 'ma351', 'ma412', 'st226', 'st321', 'st327']	[gts117, mas117]
3	am151	symbolic logic methods of proof proof of quantifier sentences real number system properties of real numbers sets properties and theorems of sets relations equivalence relations partial order relations functions and types of functions	sci	['am211', 'am457', 'ma332', 'ma337', 'ma338', 'ma366', 'ma368']	[]
4	am200	logic algebra of sets relations and functions recurrence relations generating functions basic concepts in graph theory number representation in computer boolean algebra and combinatorial circuits finite-state machines finite automata basic concepts in algebraic system note : 1 this subject is equivalent to cs 101 discrete structures 2 there is no credit for students who currently taking or have earned credits of cs 101	sci	['cs.101', 'cs.101', 'am346', 'am411', 'am448', 'am451', 'ma338', 'ma366', 'ma367']	[cn200, des201]

รูปที่ 4-1 Final Dataframe

ข้อมูลนี้สามารถนำไปใช้เป็นเครื่องมือสำคัญสำหรับการพิจารณาและวางแผนทางในการเทียบโอนหน่วยกิตภายในมหาวิทยาลัยธรรมศาสตร์ โดยเฉพาะสำหรับการดำเนินงานในอนาคตของธนาคารหน่วยกิต (Credit Bank) ซึ่งเป็นแนวคิดที่ช่วยสนับสนุนให้นักศึกษามีความยืดหยุ่นในการศึกษา สามารถโอนย้ายหน่วยกิตระหว่างคณะหรือหลักสูตรต่าง ๆ ได้อย่างมีประสิทธิภาพ

นอกจากนี้ การจัดทำข้อมูลดังกล่าวยังช่วยให้หน่วยงานที่เกี่ยวข้องสามารถนำไปพัฒนาระบบการจัดการข้อมูลหน่วยกิตที่มีความถูกต้อง โปร่งใส และตอบสนองต่อความต้องการของนักศึกษาได้ดียิ่งขึ้น อีกทั้งยังเป็นการเพิ่มโอกาสทางการศึกษาและสร้างความคล่องตัวในการเรียนรู้สำหรับนักศึกษาที่ต้องการย้ายคณะหรือเทียบโอนหลักสูตรให้สอดคล้องกับความสนใจและเป้าหมายในการเรียนรู้ของตนเอง

4.2. Website

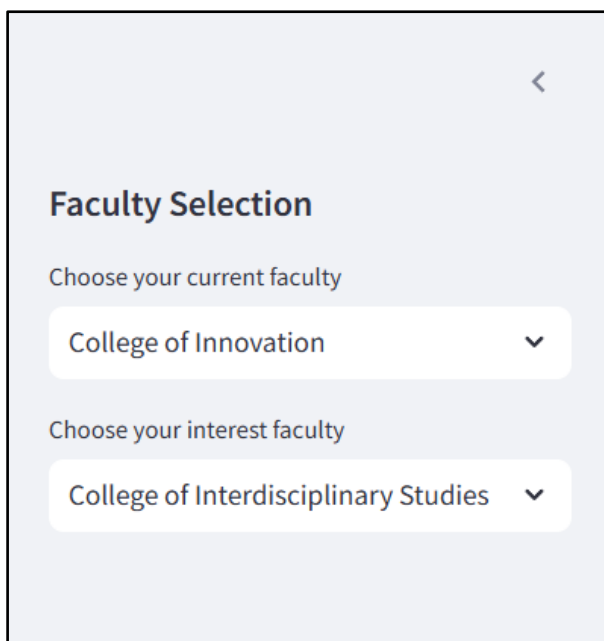
เว็บไซต์ที่สามารถศึกษาคณะที่สนใจ จากทั้งหมด 5 คณะ ของมหาวิทยาลัยธรรมศาสตร์ และสามารถเลือกคณะที่ต้องการเทียบโอนรายวิชาได้ และมี course description เพื่อประกอบการตัดสินใจ โดยเว็บไซต์ถูกออกแบบมาเพื่อเป็นแหล่งข้อมูลกลางสำหรับการศึกษารายละเอียดของคณะที่สนใจภายในมหาวิทยาลัยธรรมศาสตร์ ซึ่งครอบคลุมทั้งหมด 5 คณะ ได้แก่ คณะวิศวกรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, วิทยาลัยสหวิทยาการ, วิทยาลัยนวัตกรรม และสถาบันเทคโนโลยีนานาชาติสิรินธร(SIIT)

โดยมีวัตถุประสงค์หลักเพื่ออำนวยความสะดวกให้ผู้สนใจในการศึกษาข้อมูลคณะที่เกี่ยวข้อง และนักศึกษาในอนาคต พร้อมทั้งสามารถตรวจสอบความเป็นไปได้ในการเทียบโอนรายวิชาระหว่างคณะต่าง ๆ ภายในมหาวิทยาลัยธรรมศาสตร์ ซึ่งจะเป็นประโยชน์อย่างยิ่งต่อการตัดสินใจในกรณีที่นักศึกษามีความประสงค์จะย้ายคณะ หรือโอนย้ายหลักสูตรการศึกษาให้เหมาะสมกับความสนใจและเป้าหมายทางการเรียนรู้ของตนเอง อีกทั้งยังเป็นแนวทางในการสร้างธนาคารหน่วยกิตของมหาวิทยาลัยธรรมศาสตร์ในอนาคต

4.2.1. ฟังก์ชันการใช้งานหลักของเว็บไซต์

4.2.1.1. การแสดงผลข้อมูลคณะที่สามารถเทียบโอนได้

ผู้ใช้งานสามารถเลือก คณะปัจจุบัน และ คณะที่สนใจ ผ่านทางเมนู Side Bar ซึ่งได้รับการออกแบบมาให้ใช้งานง่าย ช่วยให้ผู้ใช้สามารถระบุคณะต้นทางและคณะปลายทางที่ต้องการทำการเทียบโอนได้อย่างสะดวกรวดเร็ว



รูปที่ 4-2 Sidebar

4.2.1.2. การแสดง Course Description ของแต่ละรายวิชา

เว็บไซต์จะแสดงข้อมูล คำอธิบายรายวิชา (Course Description) อย่างละเอียดจากทั้งคณะต้นทางและคณะปลายทาง เพื่อให้ผู้ใช้งานสามารถทำการดูรายละเอียดรายวิชาที่มีความคล้ายคลึงกันได้อย่างชัดเจน ซึ่งช่วยสนับสนุนการตัดสินใจในการเทียบโอนรายวิชาได้อย่างแม่นยำและมีประสิทธิภาพ

The screenshot displays the 'Thammasat University Credit Bank' interface. At the top, there are two buttons: 'College of Interdisciplinary Studies' (red) and 'College of Innovation' (yellow). Below these, there are two input fields for 'Current Faculty Course Code' and 'Interest Faculty Course Code'. The 'Current Faculty Course Code' field contains 'cis107' and the 'Interest Faculty Course Code' field contains 'dx310'. A double-headed arrow icon is positioned between the two fields. Below the input fields, there are two columns of course descriptions. The left column is for 'cis107 course description' and the right column is for 'dx310 course description'. Both descriptions are identical and describe a data science and innovation capability strategy for building innovation capacity barriers to innovation innovation readiness manage organization change resulting from innovation organization structure cultures and leadership for innovation sustaining innovation fundamental concepts and analytical tools that help assess opportunities and develop entrepreneurial ventures and how these concepts and tools may be applied to real world business situations and opportunities develop define and clearly communicate a new business concept or initiative in an efficient effective and compelling manner.

รูปที่ 4-3 Course Description compare

4.2.1.3. การจัดการข้อมูลผ่านระบบที่เป็นมิตรต่อผู้ใช้งาน

ข้อมูลทั้งหมดถูกนำเสนอในรูปแบบที่เข้าใจง่ายและมีความเป็นระบบ โดยอาศัยการจัดเก็บและประมวลผลข้อมูลผ่านการวิเคราะห์แล้ว พร้อมทั้งนำเสนอในลักษณะที่สามารถเข้าถึงและใช้งานได้อย่างสะดวก ผ่านหน้า Home Page ซึ่งรวบรวมข้อมูลสำคัญของคณะต่าง ๆ ไว้อย่างครบถ้วน พร้อมลิงก์เชื่อมต่อไปยังเว็บไซต์ของคณะนั้น ๆ เพื่ออำนวยความสะดวกให้แก่ผู้ใช้งาน

โดยสามารถใช้เว็บไซต์เพื่อศึกษารายละเอียดเกี่ยวกับรายวิชาของคณะทั้ง 5 คณะข้างต้น ภายในมหาวิทยาลัยธรรมศาสตร์ได้ และตรวจสอบคู่รายวิชาที่สามารถเทียบโอนกันได้ ระหว่างคณะต่าง ๆ ผ่านการแสดงผลข้อมูลที่ชัดเจนและเป็นระบบใช้ข้อมูลคำอธิบายรายวิชาเพื่อเป็น แนวทางในการประกอบการตัดสินใจ ก่อนทำการย้ายคณะหรือโอนย้ายหลักสูตร รวมทั้งสนับสนุนแนวคิดการจัดทำธนาคารหน่วยกิต (Credit Bank) สำหรับมหาวิทยาลัยธรรมศาสตร์ซึ่งเป็นเครื่องมือสำคัญในการสร้างความยืดหยุ่นและคล่องตัวในระบบการศึกษา

บทที่ 5 ความเชื่อมโยงกับวิชาต่าง ๆ ในโมดูล

5.1. วช.310 การสำรวจและการเตรียมข้อมูล

ในการดำเนินโครงการนี้ทางคณะผู้จัดทำได้นำองค์ความรู้ที่ได้จากรายวิชา วช.310 การสำรวจและการเตรียมข้อมูล มาประยุกต์ใช้ในกระบวนการสำคัญเบื้องต้น ได้แก่ การเก็บรวบรวมข้อมูล เพื่อนำไปใช้ในการวิเคราะห์และประมวลผลข้อมูลต่อไป โดยได้เลือกใช้เทคนิค Web Scraping ผ่านเครื่องมือ Selenium กระบวนการนี้มุ่งเน้นไปที่การดึงข้อมูลที่เกี่ยวข้องจากเว็บไซต์ของมหาวิทยาลัยธรรมศาสตร์ โดยเฉพาะข้อมูลรายวิชาที่เปิดสอนในคณะต่าง ๆ ซึ่งเป็นคณะเป้าหมายทั้ง 5 คณะ

5.1.1. การเก็บรวบรวมข้อมูล (Data Collection)

ใช้ Selenium ในการเขียนสคริปต์สำหรับการเข้าไปดึงข้อมูลจากเว็บ TU Reg ซึ่งเป็นแหล่งข้อมูลหลักที่รวบรวมรายละเอียดของรายวิชาทั้งหมด ข้อมูลที่ได้ประกอบไปด้วย รหัสวิชา (Course Code) คำอธิบายรายวิชา (Course Description) รายวิชาบังคับก่อน (Disciplinary Course) และ รายวิชาต่อเนื่อง (Consecutive Course)

โดยกระบวนการดังกล่าวจะถูกดำเนินการแยกตามแต่ละคณะ เพื่อให้ได้ข้อมูลที่ครบถ้วนตรงตามเป้าหมายของโครงการ โดยข้อมูลที่ได้มีดังนี้

ข้อมูลรายวิชาของแต่ละคณะ ทั้งหมด 5 คณะ 4272 ข้อมูล

	code	desc	add1	add2	add3
0	AM101	Limit and continuity of single variable functi...	[disciplinary course(s): ไม่นับหน่วยกิตให้ผิ ที่...	[consecutive course(s): AM102, ST326]	□
1	AM101	Limit and continuity of single variable functi...	[disciplinary course(s): ไม่นับหน่วยกิตให้ผิ ที่...	[consecutive course(s): AM102, ST326]	□
2	AM101	Limit and continuity of single variable functi...	[disciplinary course(s): ไม่นับหน่วยกิตให้ผิ ที่...	[consecutive course(s): AM102, ST326]	□
3	AM101	Limits and continuity of single variable funct...	[disciplinary course(s): หมายเหตุ : ไม่นับ หน่วย...	[consecutive course(s): AM102, AM201, ST326, S...	□
4	AM101	Limits and continuity of single variable funct...	[disciplinary course(s): หมายเหตุ : ไม่นับ หน่วย...	[consecutive course(s): AM102, AM201, ST326, S...	□

รูปที่ 5-1 ชุดข้อมูลของรายวิชาก่อนทำ Data Preprocessing

5.1.2. การตรวจสอบคุณภาพของข้อมูลเบื้องต้น

หลังจากทำการดึงข้อมูลเสร็จสิ้น ข้อมูลที่ได้จะถูกจัดเก็บในรูปแบบ DataFrame เพื่อให้ง่ายต่อการตรวจสอบและการนำไปใช้งาน และทำการตรวจสอบความสมบูรณ์ของข้อมูลที่รวบรวมมาได้แก่ การดูจำนวนข้อมูลซ้ำ (Duplicate Data), ข้อมูลที่ขาดหายไป (Missing Values) และความถูกต้องของโครงสร้างข้อมูล

5.1.3. การเตรียมข้อมูลก่อนประมวลผล (Data Preprocessing)

ก่อนที่จะเข้าสู่ขั้นตอนการวิเคราะห์ข้อมูล รายวิชาที่รวบรวมได้จากทั้ง 5 คณะ จะต้องผ่านการทำให้ Data Preprocessing ซึ่งเป็นขั้นตอนสำคัญในการทำความสะดวกข้อมูลให้พร้อมสำหรับการนำไปวิเคราะห์

ขั้นตอนนี้ประกอบไปด้วยการจัดรูปแบบข้อมูลให้เหมาะสม เช่น การจัดการข้อมูลที่เป็นข้อความ การลบช่องว่างที่ไม่จำเป็น (White Spaces) การตรวจสอบความถูกต้องของคำอธิบายรายวิชา และการปรับเปลี่ยนรูปแบบข้อมูลให้เป็นมาตรฐานเดียวกัน

ข้อมูลรายวิชาของแต่ละคณะ ทั้งหมด 5 คณะ หลังการทำ Data Preprocessing 1833 ข้อมูล

course code	course description	disciplinary course	consecutive course	faculty	cancel_course
19 am360, ma391	discussion on topics of current interest in applied mathematics basic skills for academic reading proposal topic regulations and procedures literature review designing and planning of the project report writing and topic of project presentation	['disciplinary course(s): am260\nnstamได้ am.260 หรือ ได้รับอนุมัติจากผู้บรรยาย']	['consecutive course(s): am460']	sci	[am260, am.260, am460]
59 at437, at438	biology of animal behaviour behavioural genetics memory and learning foraging and ingestive behaviour social and reproductive behaviour farm animal behaviour regulations and assessment of animal welfare field trips	[]	[]	sci	[]
213 cs303, cs401	senior projects will be assigned or approved by the department this course prepares students for developing a proposal for doing research writing research papers and presenting the proposal to the committee	['disciplinary course(s): cs300 and cs301 or\nncs300 and cs302\nnstสำหรับนักศึกษาตั้งแต่ ชั้นปีที่ 3 ขึ้นไป']	['consecutive course(s): cs403']	sci	[cs300, cs301, cs300, cs302, cs403]
235 cs369, cs485	basic systems and protocols for providing services on the internet developing and deploying web applications topics covered web application development techniques for both client side and server side session management interfacing with other online services web application security web design patterns and reusable web application components	['disciplinary course(s): cs234 or\nncs322']	[]	sci	[cs234, cs322]
238 cs379, cs466	deterministic and stochastic grammars morphological analysis syntax analysis semantic analysis discourse analysis and applications (eg machine translation speech recognition and synthesis and text mining)	['disciplinary course(s): cs265']	[]	sci	[cs265]

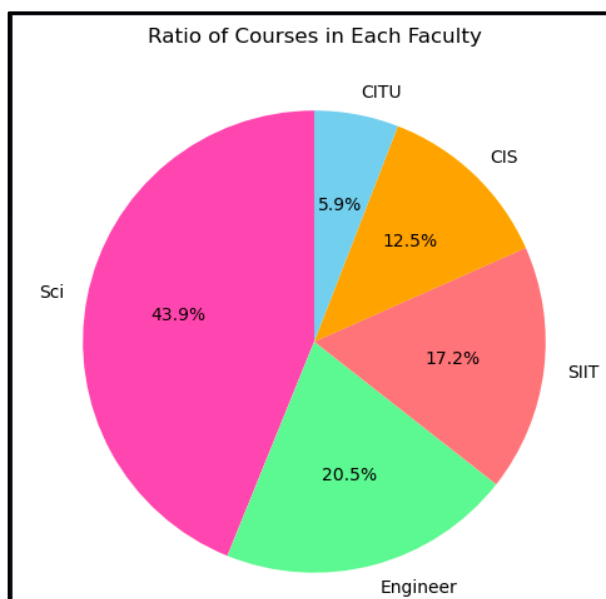
รูปที่ 5-2 ชุดข้อมูลของรายวิชาหลังจากทำ Data Preprocessing

5.2. วข.311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล

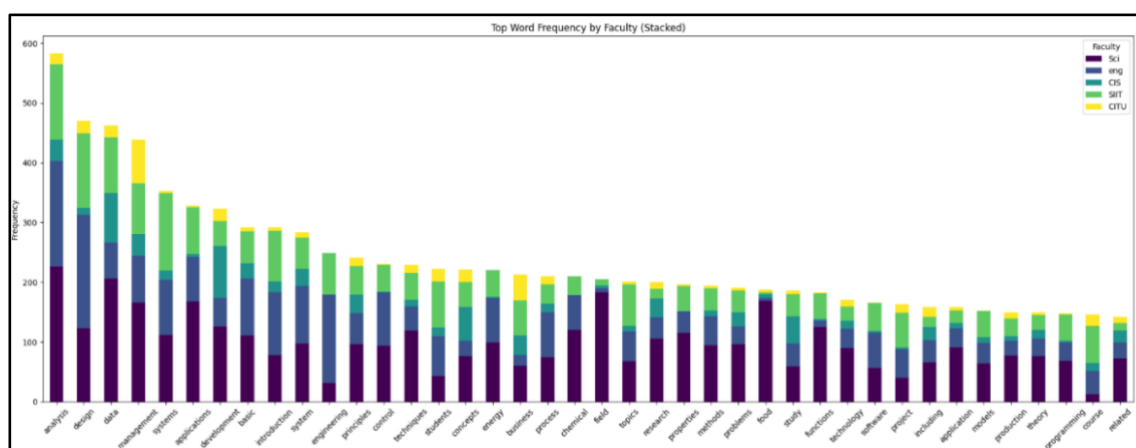
นำความรู้ในวิชา วข.311 อัลกอริทึมของวิทยาศาสตร์ข้อมูล มาประยุกต์ใช้ในการทำ NLP โดยการใช้ Transformer ซึ่งประกอบด้วย SBERT (Sentence-BERT) เป็นการเปรียบเทียบความคล้ายคลึงกันของข้อความ และการทำ Sentence Embedding แปลงข้อมูลจาก Text ไปเป็น Vector จากนั้นนำเวกเตอร์ที่ได้จากการทำ Sentence Embedding ไปคำนวณหาความคล้ายคลึงกันของข้อความโดยใช้ Cosine Similarity การกำหนดเกณฑ์และเงื่อนไขในการจับคู่รายวิชา เพื่อให้ได้คู่รายวิชาที่มีความเหมาะสม ทางคณะผู้จัดทำได้กำหนด Threshold หรือเกณฑ์ขั้นต่ำของความคล้ายคลึงกันที่ต้องการ นอกจากนี้ ยังได้กำหนดเงื่อนไขเพิ่มเติม เช่น ไม่จับคู่รายวิชาที่อยู่ภายในคณะเดียวกัน เพื่อให้ได้ผลลัพธ์ที่เป็นไปตามวัตถุประสงค์ของโครงการ ซึ่งมุ่งเน้นไปที่การเทียบโอนรายวิชาระหว่างคณะ

5.3. วข.312 ระบบธุรกิจอัจฉริยะ

นำความรู้ในวิชา วข.312 ระบบธุรกิจอัจฉริยะ มาประยุกต์ใช้ในการทำ Data Management เพื่อช่วยในการจัดการกับข้อมูลให้ข้อมูลมีคุณภาพ เชื่อถือได้ มีความปลอดภัย และพร้อมใช้งาน ซึ่งจะช่วยเสริมการวิเคราะห์ได้อย่างมีประสิทธิภาพ และใช้ Visual Analytics (EDA) เพื่อทำให้สามารถทำความเข้าใจ และช่วยให้สามารถมองเห็นภาพรวมของข้อมูลได้ดีขึ้นผ่านกราฟ ดังเช่น pie chart แสดงจำนวนรายวิชาของแต่ละคณะ จำนวนทั้งหมด 5 คณะ จาก 1833 รายวิชา และกราฟแท่งแบบต่อที่แสดงค่าความถี่ของจำนวนคำที่พบบ่อยที่สุดในแต่ละคณะ



รูปที่ 5-3 pie chart แสดงจำนวนรายวิชาทั้งหมดของ 5 คณะ (คิดเป็น %)



รูปที่ 5-4 กราฟแท่งแบบต่อแสดงค่าความถี่ของคำที่พบบ่อยที่สุดในแต่ละคณะ

5.4. วข.313 การวิเคราะห์การตลาด

นำความรู้ในวิชา วข.313 การวิเคราะห์การตลาดมาประยุกต์ใช้โดยมีการเก็บรวบรวมข้อมูลในรูปแบบข้อมูลทุติยภูมิโดยใช้วิธีการดึงข้อมูลจากเว็บไซต์ TU REG และมีการใช้การวิจัยทางการตลาด ในการวิจัยผลิตภัณฑ์ซึ่งในที่นี้คือการวิเคราะห์หลักสูตรการศึกษาระดับปริญญาตรีภายในมหาวิทยาลัยธรรมศาสตร์ ภาคการศึกษาที่ 1/2566 และ 2/2566 ภายใต้ 5 คนที่เรามุ่งเน้น เพื่อหาข้อเท็จจริงต่าง ๆ ที่เป็นประโยชน์ต่อทั้งมหาวิทยาลัย อาจารย์ และนักศึกษา และการทำให้ Insight เป็น Actionable Insight เพื่อให้สามารถใช้เพื่อช่วยในการเทียบโอนรายวิชาและหน่วยกิต รวมไปถึงการทำธนาคาร หน่วยกิต (Credit Bank)

บทที่ 6 บทสรุป

จากการวิเคราะห์ความคล้ายคลึงของรายวิชาในทั้งหมด 5 คณะ ของมหาวิทยาลัยธรรมศาสตร์ ระดับปริญญาตรี ได้แก่ คณะวิศวกรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, วิทยาลัยสหวิทยาการ, วิทยาลัยนวัตกรรม และสถาบันเทคโนโลยีนานาชาติสิรินธร (SIIT) เพื่อนำข้อมูลที่ได้จากการวิเคราะห์ความคล้ายคลึงกันของคำอธิบายรายวิชาต่าง ๆ ภายในคณะที่กล่าวไปข้างต้น ไปใช้เพื่อหารายวิชาที่สามารถเทียบโอนกันได้ โดยจะใช้ Selenium ในการดึงข้อมูลจากเว็บไซต์ และใช้ SBERT model ซึ่งเป็น Transformers-based model ในกระบวนการวิเคราะห์คำอธิบายรายวิชาพร้อมแปลงคำเป็น vector และใช้เทคนิค Cosine Similarity ในการเทียบหาความสอดคล้องของรายวิชาต่าง ๆ ซึ่งผลลัพธ์ที่ได้จะช่วยให้ทราบว่ามียรายวิชาใดบ้างที่มีความสอดคล้องกัน

การวิเคราะห์ การเลือกสร้างโมเดลและการเลือกเทคนิคที่มีประสิทธิภาพ สามารถสร้างประโยชน์ได้หลากหลายด้าน โดยทางมหาวิทยาลัยธรรมศาสตร์สามารถนำผลลัพธ์ของการวิเคราะห์ไปสนับสนุนการเทียบโอนรายวิชาและหน่วยกิต พร้อมทั้งใช้เป็นแนวทางในการสร้างระบบธนาคารหน่วยกิต (Credit Bank) ของมหาวิทยาลัยได้ โดยการนำข้อมูลที่ได้จากการวิเคราะห์ความคล้ายคลึงของรายวิชาต่าง ๆ นั้น มาดูว่ามีรายวิชาใดที่มีความสอดคล้องกัน ซึ่งจะทำให้ทราบได้ว่ามียรายวิชาใดบ้างที่สามารถเทียบโอนกันได้

ดังนั้นผลลัพธ์จากการวิเคราะห์นี้ไม่เพียงแต่ช่วยสนับสนุนการทำงานของหน่วยงานที่เกี่ยวข้อง แต่ยังเป็นการวางรากฐานที่สำคัญในการพัฒนาระบบธนาคารหน่วยกิต (Credit Bank) ของมหาวิทยาลัยธรรมศาสตร์ให้มีความครอบคลุมและทันสมัย โดยข้อมูลที่ได้จากการวิเคราะห์นี้สามารถต่อยอดเพื่อพัฒนาระบบสารสนเทศหรือเครื่องมือดิจิทัลที่จะช่วยเพิ่มประสิทธิภาพในการตรวจสอบและเทียบโอนรายวิชาได้อย่างเป็นระบบในอนาคต

บรรณานุกรม

- กระทรวงศึกษาธิการ. (2566). คู่มือดำเนินงานระบบธนาคารหน่วยกิตแห่งชาติ. สืบค้นจาก <https://backoffice.onec.go.th/uploads/Book/2061-file.pdf>
- กระทรวงศึกษาธิการ. (2567). ประกาศกระทรวงศึกษาธิการ เรื่อง แนวทางการดำเนินงานธนาคารหน่วยกิตระดับการศึกษาขั้นพื้นฐานและการอาชีวศึกษา พ.ศ.2567. สืบค้นจาก <https://ratchakitcha.soc.go.th/documents/27701.pdf>
- จรรยา แหนมเจริญ, ยงยุทธ ช่อจำปา, สุรพล คล้ายภมร และสายยาใจ ฟังประชาชน. (2564). ระบบเทียบโอนรายวิชาและผลการเรียนด้วยการทำเหมืองข้อความ. สืบค้นจาก <https://e-research.siam.edu/wp-content/uploads/2021/02/Proceeding-Computer-Science-by-Janya-Yamcharoen-KRUNAC2020.pdf>
- ปริญญา ทองคำ. (2562). การพัฒนาระบบบริหารจัดการงานเทียบโอนรายวิชาเพื่อเพิ่มประสิทธิภาพการทำงานในสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏมหาสารคาม. สืบค้นจาก <https://ph02.tci-thaijo.org/index.php/project-journal/article/view/184520>
- พนัสดา สิริรังศรี, รัตติกร โสมสมบัติ, อรุณี ตระการไพโรจน์, นลินี รักธรรม ณ นคร และธนวุฒิ แก้วนุช. (ม.ป.ป.). ระบบการสะสมหน่วยการเรียนรู้ หรือ ธนาคารหน่วยกิต (Credit Bank System). สืบค้นจาก <https://www.dpu.ac.th/ces/upload/km/1494553177.%E0%B8%94%E0%B8%A3>
- สำนักงานสภานโยบายอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรมแห่งชาติ. (2022). ธนาคารหน่วยกิตแห่งชาติโอกาสทางการศึกษา สำหรับทุกช่วงวัย. สืบค้นจาก <https://www.nxpo.or.th/th/11862/>
- Anirban Sen. (2023). *SBERT: How to Use Sentence Embeddings to Solve Real-World Problems*. Retrieved from <https://anirbansen2709.medium.com/sbert-how-to-use-sentence-embeddings-to-solve-real-world-problems-f950aa300c72>
- Pairode Jaroensri & Weerapat Satitkanitkul. (2567). *Transformer – โครงสร้างพื้นฐานของ Generative AI*. สืบค้นจาก <https://bdi.or.th/big-data-101/transformer-generativeai/Lo/156d217b0a911af97fa1b5a71dc909ccef7a8028#paper-topics>
- The White Marketing. (ม.ป.ป.). รู้จัก BERT AI อัลกอริทึมใหม่ของกูเกิ้ล. สืบค้นจาก

<https://thewhitemarketing.com/bert-algorithm/>