การบ้านที่ 1 (1/2564)

รายวิชา คพ.348 แบบจำลองสำหรับวิทยาการข้อมูล

ข้อที่ 1

1.1. ให้นักศึกษาอธิบายความหมายของ Data Science ตามความเข้าใจของนักศึกษา

Data Science หรือ วิทยาศาสตร์ ข้อมูล หรือ วิทยาการข้อมูล หมายถึง การศึกษาข้อมูลและวิเคราะห์ข้อมูลเหล่านั้นให้เกิดประโยชน์ ซึ่งรวมหลาย ๆ เทคนิคทางคณิตศาสตร์ในการวิเคราะห์ ตัวอย่างของ Data Science เช่น แอปพลิเคชันอ่านนิยาย ตัวแอปจะเก็บข้อมูลหนังสือที่ผู้ใช้อ่านและทำการวิเคราะห์ความชอบ จากนั้นแอปจะแนะนำหนังสือที่อาจตรงกับแนวที่ผู้ใช้ชื่อชอบ เป็นต้น จากตัวอย่างจะเห็นได้ว่าแอปพลิเคชันจะศึกษาข้อมูลของผู้ใช้ จากนั้นจะทำการวิเคราะห์ข้อมูลเหล่านั้นออกมาเพื่อแนะนำหนังสือที่ผู้ใช้ชอบซึ่งถือเป็นการทำให้เกิดประโยชน์

1.2. ให้นักศึกษาเข้าไปที่เว็บไซต์ <https://data.go.th/> (หรือแหล่งข้อมูลอื่นที่นักศึกษาสนใจ)

1.2.1. เลือกข้อมูลที่นักศึกษาสนใจมา 3 แบบ

1.จำนวนนักเรียนแยกตามสัญชาติ ปีการศึกษา 2564 เทอม 1 จาก กระทรวงศึกษาธิการ

ลิงก์ : <https://data.go.th/dataset/nation2564-1>

2.นักวิจัยดำเนินการด้าน Zero Waste จาก กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม

ลิงก์ : <https://data.go.th/dataset/researcherzerowaste>

3.ปริมาณการใช้งานเน็ตประชารัฐ จาก กระทรวงดิจิทัลเพื่อเศษรฐกิจและสังคม

ลิงก์ : <https://data.go.th/dataset/village_internet_usage>

1.2.2. อธิบายรายละเอียดเกี่ยวกับข้อมูลที่เลือกมาแต่ละแบบ และระบุ meta-data

1.ข้อมูลขื่อ “จำนวนนักเรียนแยกตามสัญชาติ ปีการศึกษา 2564 เทอม 1” มาจากองค์กร “สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 4” เป็นข้อมูลด้านการศึกษา ข้อมูลเกี่ยวกับจำนวนนักเรียนที่แยกตามสัญชาติ ปีการศึกษา 2564 เทอม 1 ในเขตปทุมธานี

Meta-data ของข้อมูลชุดที่ 1 มีดังนี้

|  |  |
| --- | --- |
| **รายการคำอธิบายชุดข้อมูล** | **คำอธิบายข้อมูล** |
| ชื่อชุดข้อมูล | จำนวนนักเรียนแยกตามสัญชาติ ปีการศึกษา 2564 เทอม 1 |
| องค์กร | สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 4 |
| ชื่อผู้ติดต่อ | วันเพ็ญ แตงงาม |
| อีเมลผู้ติดต่อ | ict@ssps4.go.th |
| คำสำคัญ | นักเรียน, ปทุมธานี, สัญชาติ, โรงเรียน |
| รายละเอียด | จำนวนนักเรียนแยกตามสัญชาติ ปีการศึกษา 2564 เทอม 1 ในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาปทุมธานี |
| ขอบเขตเชิงภูมิศาสตร์ | ปทุมธานี |
| วันที่สร้างชุดข้อมูล | 27 สิงหาคม 2564 |
| กลุ่มชุดข้อมูล | การศึกษา |
| ความถี่ในการปรับปรุง | รายไตรมาส |
| วันที่ปรับปรุงข้อมูลล่าสุด | 27 สิงหาคม 2564 |
| สัญญาอนุญาต | DGA Open Government License |

2.ข้อมูลขื่อ “นักวิจัยดำเนินการด้าน Zero Waste” มาจากองค์กร “สำนักงานการวิจัยแห่งชาติ” เป็นข้อมูลด้านวิทยาศาสตร์ เทคโนโลยีดิจิทัล และนวัตกรรม ข้อมูลเกี่ยวกับนักวิจัยที่ดำเนินการด้าน Zero Waste (ขยะเหลือศูนย์) จากระบบข้อมูลสารสนเทศวิจัยและนวัตกรรมแห่งชาติ

Meta-data ของข้อมูลชุดที่ 2 มีดังนี้

|  |  |
| --- | --- |
| **รายการคำอธิบายชุดข้อมูล** | **คำอธิบายข้อมูล** |
| ชื่อชุดข้อมูล | นักวิจัยดำเนินการด้าน Zero Waste |
| องค์กร | สำนักงานการวิจัยแห่งชาติ |
| ชื่อผู้ติดต่อ | นางสาวรัตนา สุวรรณวิชนีย์ |
| อีเมลผู้ติดต่อ | rattana.s@nrct.go.th |
| คำสำคัญ | Zero Waste, นักวิจัย |
| รายละเอียด | นักวิจัยดำเนินการด้าน Zero Waste จากข้อมูลในระบบข้อมูลสารสนเทศวิจัยและนวัตกรรมแห่งชาติ (ระบบ NRIIS) |
| ขอบเขตเชิงภูมิศาสตร์ | Thailand |
| วันที่สร้างชุดข้อมูล | 9 พฤศจิกายน 2563 |
| กลุ่มชุดข้อมูล | วิทยาศาสตร์ เทคโนโลยีดิจิทัล และนวัตกรรม |
| ความถี่ในการปรับปรุง | รายไตรมาส |
| วันที่ปรับปรุงข้อมูลล่าสุด | 7 เมษายน 2564 |
| สัญญาอนุญาต | DGA Open Government License |

3.ข้อมูลขื่อ “ปริมาณการใช้งานเน็ตประชารัฐ” มาจากองค์กร “สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม” เป็นข้อมูลด้านวิทยาศาสตร์ เทคโนโลยีดิจิทัล และนวัตกรรม ข้อมูลเกี่ยวกับปริมาณการใช้งานอินเทอร์เน็ตของโครงการเน็ตประชารัฐ

Meta-data ของข้อมูลชุดที่ 3 มีดังนี้

|  |  |
| --- | --- |
| **รายการคำอธิบายชุดข้อมูล** | **คำอธิบายข้อมูล** |
| ชื่อชุดข้อมูล | ปริมาณการใช้งานเน็ตประชารัฐ |
| องค์กร | สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม |
| ชื่อผู้ติดต่อ | ศูนย์ขับเคลื่อนการพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม |
| อีเมลผู้ติดต่อ | wittawat.c@mdes.go.th |
| คำสำคัญ | ปริมาณการใช้งาน, เน็ตประชารัฐ |
| รายละเอียด | ชุดข้อมูลปริมาณการใช้งานเน็ตประชารัฐของหมู่บ้าน จำนวน ๒๔,๗๐๐ หมู่บ้าน รายเดือน |
| ขอบเขตเชิงภูมิศาสตร์ | Thailand |
| วันที่สร้างชุดข้อมูล | 4 มกราคม 2564 |
| กลุ่มชุดข้อมูล | วิทยาศาสตร์ เทคโนโลยีดิจิทัล และนวัตกรรม |
| ความถี่ในการปรับปรุง | รายเดือน |
| วันที่ปรับปรุงข้อมูลล่าสุด | 19 กรกฎาคม 2564 |
| สัญญาอนุญาต | DGA Open Government License |

1.2.3. ตั้งคำถามที่น่าสนใจเกี่ยวกับข้อมูลแต่ละแบบที่เลือกมา

คำถาม ข้อมูลชุดที่ 1

1.จำนวนนักเรียนสัญชาติไทยในระดับชั้นมัธยมศึกษาปีที่ 6 มีเท่าไหร่

2.จำนวนความแตกต่างระหว่างนักเรียนสัญชาติญี่ปุ่นกับสัญชาติจีนในระดับชั้นมัธยมศึกษาปีที่ 6 เป็นเท่าไหร่

3.แต่ละโรงเรียน มีจำนวนนักเรียนต่างชาติเป็นเท่าไหร่ และโรงเรียนอะไรที่มีจำนวนนักเรียนต่างชาติมากที่สุดและน้อยที่สุด

คำถาม ข้อมูลชุดที่ 2

1.จำนวนงานวิจัยที่มีนักวิจัยตำแหน่งทางวิชาการเป็นศาสตราจารย์และอยู่มหาวิทยาลัยธรรมศาสตร์

2.รายชื่องานวิจัยที่มีสิทธิบัตรและมีผลงานตีพิมพ์

3.ในปี 2562 จงระบุงานวิจัยที่มีการจัดการกับพลาสติกไม่ทางตรงก็ทางอ้อม

คำถาม ข้อมูลชุดที่ 3

1.จังหวัดใดที่มีจำนวนผู้ใช้งานเน็ตประชารัฐมากที่สุด

2.ภาคไหนในไทยที่มีปริมาณการดาวน์โหลดเยอะที่สุด

3.โดยเฉลี่ยแล้วคนที่ใช้งานเน็ตประชารัฐส่วนใหญ่ใช้อุปกรณ์กันกี่เครี่อง

ข้อที่ 2

2.1. อธิบายความหมายของ แบบจำลองความน่าจะเป็น (Probability Model) ตามที่นักศึกษาเข้าใจ

แบบจำลองความน่าจะเป็น คือ การอธิบายสถานการณ์ที่ไม่แน่นอน โดยต้องมี 2 องค์ประกอบ 1.Sample Space คือ set ของผลลัพธ์ทั้งหมดที่เกิดจากการทดลองที่เราสนใจ และ 2.Probability Law คือ กฎความน่าจะเป็น เขียนด้วย P(A) ยกตัวอย่าง เช่น ลูกโลก ก็ทำการกำหนด Probability Law ลงบน Sample Space โดยเหตุการณ์ที่เราสนใจก็จะเป็นที่ตั้งของแต่ละประเทศ สภาพภูมิอากาศ เป็นต้น