

สรุปบทเรียนจากการพัฒนาความรู้ รอบที่ ๑ ประจำปี ๒๕๖๘

ของ นางสาววรยา สุธรรมชัย

เรื่อง การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล (Data Analytics) คือ กระบวนการที่ค้นหาพิจารณาความหมาย หรือรูปแบบของข้อมูลให้เป็นประโยชน์ โดยจะเริ่มจากขั้นตอนการระบุปัญหา กำหนดวัตถุประสงค์ เก็บรวบรวมข้อมูล เตรียมข้อมูลเบื้องต้น วิเคราะห์ข้อมูล และตีความผลลัพธ์ข้อมูลว่าสามารถตอบโจทย์ปัญหาได้หรือไม่ ถ้าไม่สามารถตอบโจทย์ปัญหาได้ อาจจะต้องทำการเก็บข้อมูลเพิ่มเติม

โปรแกรม RapidMiner เป็นเครื่องมือสำหรับการวิเคราะห์ข้อมูล และการสร้างโมเดลการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้งานง่าย และไม่จำเป็นต้องมีทักษะการเขียนโปรแกรมขั้นสูง สามารถใช้ในการวิเคราะห์ข้อมูลจากแหล่งต่าง ๆ เช่น ฐานข้อมูล Excel CSV SQL และอื่น ๆ

ขั้นตอนการใช้งาน RapidMiner มีดังนี้

๑. ติดตั้งโปรแกรม ดาวน์โหลดและติดตั้งโปรแกรมจากเว็บไซต์ของ RapidMiner
๒. สร้างโปรเจกต์ใหม่ เมื่อเปิดโปรแกรมขึ้นมา จะมีตัวเลือกให้สร้างโปรเจกต์ใหม่ เลือก "New Project" และตั้งชื่อโปรเจกต์
๓. โหลดข้อมูล (Import Data) เลือกแหล่งข้อมูลที่ต้องการนำเข้า โดยการใช้เครื่องมือใน RapidMiner เช่น "Read Excel" หรือ "Read CSV" เพื่อโหลดข้อมูลที่ต้องการวิเคราะห์
๔. การทำความสะอาดข้อมูล (Data Preprocessing) เพื่อเตรียมข้อมูลก่อนนำไปใช้ โดยใช้เครื่องมือต่าง ๆ ใน RapidMiner เช่น
 - การจัดการค่าผิดปกติ โดยใช้เครื่องมือ "Replace Missing Values"
 - การลบข้อมูลที่ซ้ำซ้อน โดยใช้เครื่องมือ "Remove Duplicate"
 - การจัดการข้อมูลที่มีค่าสุดโต่ง โดยใช้เครื่องมือ "Detect Outlier Distances"
 - การเลือกคุณลักษณะ โดยใช้เครื่องมือ "Select Attributes"
 - การแปลงค่าตัวแปรให้เป็นค่ามาตรฐาน โดยใช้เครื่องมือ "Normalization"
๕. การสร้างโมเดล ใช้เครื่องมือ Machine Learning ใน RapidMiner เช่น "Decision Tree" หรือต้นไม้ตัดสินใจ โดยใช้คำสั่ง Decision Tree เพื่อสร้างโมเดลการทำนาย หรือการจำแนกประเภท
๖. การประเมินผลโมเดล หลังจากสร้างโมเดลแล้ว ใช้เครื่องมือประเมินผลเพื่อวัดประสิทธิภาพของโมเดลที่สร้างขึ้น เช่น "Performance Evaluation" เพื่อดูผลการทำนายและตรวจสอบความแม่นยำของโมเดล
๗. การส่งออกผลลัพธ์ เมื่อได้ผลลัพธ์ที่ต้องการแล้ว สามารถส่งออกผลลัพธ์ในรูปแบบต่าง ๆ เช่น Excel CSV หรือกราฟโดยใช้เครื่องมือ "Write Excel" หรือ "Write CSV" เพื่อใช้ในการรายงาน
๘. การสร้างรายงาน โปรแกรม RapidMiner สามารถสร้างรายงานในรูปแบบต่าง ๆ โดยใช้เครื่องมือ "RapidMiner Studio" ซึ่งสามารถแทรกกราฟ ผลการวิเคราะห์ และข้อมูลที่สำคัญเพื่อใช้ประกอบในรายงาน

ข้อดีของการใช้ RapidMiner

๑. เป็นเครื่องมือที่มีประสิทธิภาพในการวิเคราะห์ข้อมูล และการสร้างโมเดล Machine Learning โดยไม่จำเป็นต้องเขียนโปรแกรม
๒. ใช้งานได้ง่ายด้วยอินเทอร์เฟซแบบ Drag-and-Drop
๓. มีเครื่องมือหลากหลายสำหรับการประมวลผลข้อมูล และการวิเคราะห์เชิงลึก จึงสามารถนำไปใช้วิเคราะห์ข้อมูลในงานวิจัย
๔. สามารถส่งออกข้อมูลในรูปแบบต่าง ๆ เพื่อใช้ประกอบรายงาน จึงทำให้การทำงานสะดวก และรวดเร็วยิ่งขึ้น