

สรุปบทเรียนจากการ พัฒนาความรู้

สอบ 1
ปีงบประมาณ พ.ศ. 2568

หลักสูตร
การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล
หมวดทักษะดิจิทัล
เนื้อหาวิชาโดย ผศ. ดร. สิริินดา พลหาญ

จัดทำโดย นายชัยภัทร ป้องที่
ตำแหน่ง นักวิชาการคอมพิวเตอร์ปฏิบัติการ

บทเรียนการใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล เป็นบทเรียนที่เน้นพื้นฐานของการวิเคราะห์ข้อมูล การทำ Data Science และ Machine Learning โดยใช้โปรแกรมวิเคราะห์ข้อมูล RapidMiner ขั้นตอนการวิเคราะห์ข้อมูลจะประกอบไปด้วย

๑. กำหนดคำถามและตั้งสมมติฐาน

การกำหนดคำถามและตั้งสมมติฐานในสิ่งที่ต้องการหาคำตอบจะช่วยคัดกรองข้อมูล เพื่อให้ได้เฉพาะข้อมูลที่ต้องการ

๒. กำหนดชุดข้อมูล

การกำหนดชุดข้อมูลจะช่วยให้รู้ว่าองค์กรต้องการข้อมูลแบบไหนมาวิเคราะห์ และต้องการผลลัพธ์ที่เกี่ยวข้องอย่างไรกับข้อมูล จะทำให้สามารถรวบรวม จัดเก็บ และวางโครงสร้างข้อมูลได้อย่างเป็นระบบ

๓. การรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลจะทำการดึงข้อมูลจากแหล่งเดียว หรือแหล่งอื่น ๆ ที่เกี่ยวข้อง โดยใช้เครื่องมือ เช่น คอมพิวเตอร์ แบบสอบถาม บุคลากรหรือผู้เชี่ยวชาญ เป็นต้น

๔. การทำความสะอาดข้อมูล

เพื่อให้ข้อมูลเป็นระเบียบ มีความถูกต้อง ลบข้อมูลที่มีความซ้ำซ้อน การเติมข้อมูลที่ขาดหาย โดยวิธีการต่างๆ เพื่อให้ข้อมูลมีคุณภาพและพร้อมใช้งาน

๕. การวิเคราะห์ข้อมูลเชิงตรวจสอบ

เป็นกระบวนการตรวจสอบและสำรวจข้อมูลเบื้องต้น โดยวิเคราะห์ข้อมูลที่จำเป็นก่อนที่จะเข้าสู่การวิเคราะห์เชิงลึก ซึ่งช่วยตรวจสอบความถูกต้องของชุดข้อมูลนั้น เมื่อจบกระบวนการนี้อาจจะมีการรวบรวมข้อมูลเพิ่มเติมหรือการทำความสะอาดข้อมูลมากขึ้น

๖. การทำนายและการสร้างแบบจำลองทางสถิติ

เป็นขั้นตอนของการ Train Model และ Test Model จากข้อมูลที่ได้ผ่านกระบวนการก่อนหน้านี้ เพื่อให้ได้ Model ที่สมบูรณ์ มีความถูกต้อง และผ่านการวัดประสิทธิภาพ ด้วยวิธีการต่างๆ เช่น Accuracy, Precision, และ Recall เป็นต้น

๗. การตีความผลลัพธ์ของการทำนาย

การวิเคราะห์และ Machine Learning Model โดยใช้ทักษะด้านวิทยาการข้อมูล ที่เกี่ยวกับตัวเลขและสถิติ การใช้โปรแกรมและเทคนิคขั้นสูง เพื่อให้สามารถแปลผลและอธิบายได้ว่าโมเดลทำนายได้ถูกต้องหรือไม่ และเหตุใดโมเดลถึงทำการตัดสินใจเช่นนั้น การใช้โปรแกรมและเทคนิคขั้นสูงจะช่วยให้สามารถปรับปรุงและพัฒนาโมเดลให้มีความแม่นยำมากขึ้น เช่น การใช้ Feature Importance หรือ Confusion Matrix

โดยในบทเรียนนี้ การใช้อัลกอริทึม Decision Tree ในการเรียนรู้ ซึ่งเป็นอัลกอริทึมสำหรับ Classification หรือการจำแนกกลุ่มข้อมูลออกเป็นกลุ่มต่าง ๆ ตามลักษณะหรือคุณสมบัติที่กำหนดไว้ โดยการเรียนรู้ในลักษณะนี้จะเป็น Supervised Learning หรือการเรียนรู้แบบมีผู้สอน ซึ่งหมายถึง การฝึกฝนโมเดลโดยใช้ข้อมูลที่มีป้ายกำกับ (Labeled Data) ที่มีการระบุผลลัพธ์ที่ต้องการให้ชัดเจน

ตัวอย่างการเรียนรู้จากบทเรียน

ข้อมูลที่ได้รับการทำความสะอาดและเตรียมพร้อมแล้ว ซึ่งเป็นข้อมูลของลูกค้าที่ทำการซื้อสินค้าจากร้านค้า จะถูกนำเข้าในรูปแบบไฟล์ข้อมูล (เช่น CSV หรือ Excel) โดยมีป้ายกำกับข้อมูล (Label) คือ การกลับมาใช้บริการอีกครั้งและไม่ได้กลับมาใช้บริการ เพื่อผ่านกระบวนการวิเคราะห์ข้อมูล ต่อจากนั้นทำการเลือกตัวแปร (Attributes) ที่จะใช้ในการทำนาย โดยพิจารณาตัวแปรที่มีความเกี่ยวข้องกับพฤติกรรมการซื้อสินค้าของลูกค้าและความน่าจะเป็นที่ลูกค้าจะกลับมาใช้บริการกับร้านค้า

ในขั้นตอนถัดไป จะเลือกอัลกอริทึมที่เหมาะสมสำหรับการทำนาย ซึ่งในกรณีนี้เลือกใช้อัลกอริทึม Decision Tree เนื่องจากสามารถใช้ในการจำแนกกลุ่มข้อมูลได้อย่างมีประสิทธิภาพ จากนั้นทำการแบ่งชุดข้อมูลออกเป็นสองส่วน ได้แก่ ชุดข้อมูลฝึก (Training Data) และชุดข้อมูลทดสอบ (Test Data)

หลังจากนั้นทำการฝึกฝนโมเดลด้วยข้อมูลฝึก (Training Data) และเชื่อมโยงกับอัลกอริทึม Decision Tree เพื่อสร้างโมเดลที่สามารถทำนายผลได้ จากนั้นใช้ชุดข้อมูลทดสอบ (Test Data) ในการทดสอบความแม่นยำของโมเดล

ขั้นตอนสุดท้าย คือ การตีความผลลัพธ์ที่ได้จากการทำนายเพื่อประเมินความน่าจะเป็นที่ลูกค้าจะกลับมาใช้บริการกับร้านค้าอีกครั้ง โดยการจัดกลุ่มข้อมูลตามลักษณะและคุณสมบัติที่เกี่ยวข้อง เพื่อใช้ผลลัพธ์ในการพัฒนากลยุทธ์การตลาดหรือการปรับปรุงบริการให้ตรงตามความต้องการของลูกค้า

ข้อสรุปและแนวคิดในการประยุกต์ใช้เพื่อพัฒนาองค์กร

การใช้โปรแกรม RapidMiner ในการสร้างโมเดลเพื่อวิเคราะห์ข้อมูลเชิงลึกสามารถนำไปประยุกต์ใช้ในหลายด้านเพื่อพัฒนาองค์กรได้ โดยการนำ Dataset ที่ต้องการมาทำการวิเคราะห์เพื่อทำนายหรือจำแนกประเภทในสิ่งที่เราต้องการ ยกตัวอย่างเช่น การนำไปประยุกต์ใช้ในระบบการตรวจจับการบุกรุก (IDS: Intrusion Detection System) โดยการวิเคราะห์ Log File ที่ผิดปกติจากอุปกรณ์เครือข่ายต่าง ๆ เช่น Firewall, Router, และ Switch โดยการฝึกฝนโมเดลจากชุดข้อมูล (Dataset) ของภัยคุกคาม

ก่อนที่โมเดลจะถูกนำไปใช้งานจริง จำเป็นต้องทำการประเมินประสิทธิภาพของโมเดล เช่น Accuracy, Precision, และ Recall เพื่อให้สามารถระบุประเภทของการโจมตีได้อย่างรวดเร็วและแม่นยำ รวมทั้งสามารถตอบสนองต่อเหตุการณ์ได้อย่างทันท่วงที การประยุกต์ใช้โมเดลนี้จะช่วยเพิ่มประสิทธิภาพในการตรวจจับภัยคุกคามและยกระดับประสิทธิภาพของภารกิจของหน่วยงาน

โดยสรุปแล้ว การสร้างโมเดลใน RapidMiner เป็นกระบวนการที่สะดวกและมีประสิทธิภาพ เนื่องจากช่วยลดภาระในการเขียนโปรแกรมสำหรับการวิเคราะห์หรือทำนายข้อมูล ทำให้กระบวนการพัฒนาและการนำโมเดลไปใช้งานต่าง ๆ เป็นไปอย่างรวดเร็วและมีประสิทธิภาพยิ่งขึ้น โดยเฉพาะในการนำไปประยุกต์ใช้ในการตัดสินใจและการพัฒนาระบบต่าง ๆ ที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เช่น ระบบตรวจจับภัยคุกคาม การทำนายพฤติกรรมหรือการประมวลผลข้อมูลเชิงลึกในหลายๆ ด้าน การใช้ RapidMiner ยังช่วยให้องค์กรสามารถใช้ทรัพยากรได้อย่างคุ้มค่ามากขึ้น ลดขั้นตอนที่ซับซ้อนในการพัฒนาโมเดล อีกทั้งยังสามารถช่วยพัฒนาองค์กรให้มีประสิทธิภาพและความทันสมัย ตอบสนองต่อความต้องการในยุคของการใช้ปัญญาประดิษฐ์ (AI)

แหล่งที่มา

หลักสูตร : การใช้โปรแกรมดิจิทัลเพื่อการวิเคราะห์ข้อมูล

ด้านการพัฒนา (ระบุ ✓) : ☒ ดิจิทัล

☐ ความรู้เฉพาะด้านเพื่อใช้ในการปฏิบัติงาน

☐ เพิ่มศักยภาพตนเอง/ประสิทธิภาพในการทำงาน

บรรยายโดย : ชื่อ-สกุล ผศ. ดร. สิริندا พลหาญ ตำแหน่ง อาจารย์คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยหอการค้าไทย

หน่วยงานผู้รับผิดชอบ : กรมพัฒนาที่ดิน

วิธีการพัฒนาตนเอง : OCSC Learning Portal ศูนย์การเรียนรู้ทางสื่ออิเล็กทรอนิกส์แบบบูรณาการ
สำนักงาน ก.พ.

วันที่ได้รับการฝึกอบรม : 20 ก.พ. 2568 **สถานที่ :** กรมพัฒนาที่ดิน