

Practical Data Mining Analytics with RapidMiner Studio 7

ระยะเวลาอบรม 3 วัน

โลกในยุคปัจจุบันได้ก้าวเข้าไปสู่ยุคที่เรียกว่า "Big Data" หรือ "ข้อมูลอภิมหาศาล" เนื่องจากในแต่ละวันมีข้อมูลเกิดขึ้นมากมาย อาทิเช่น ข้อมูลสมาชิกของ Facebook ข้อมูลการซื้อสินค้าจากในซูเปอร์มาร์เกตต่างๆ และเพื่อให้เกิดประโยชน์มากที่สุดเราจำเป็นต้องนำข้อมูลอภิมหาศาลเหล่านี้มาทำการวิเคราะห์ (analyze) ซึ่งเทคนิคหนึ่งที่ได้รับการยอมรับอย่างสูงในปัจจุบัน คือ เทคนิค Data Mining ซึ่งเป็นเทคนิคที่ค้นหาความสัมพันธ์ในข้อมูล เช่น ถ้าลูกค้าซื้อเบียร์แล้วลูกค้าจะซื้อผ้าอ้อมร่วมไปด้วย หรือถ้าเรากด Like หน้า Facebook page เราจะเห็นว่า Facebook มีระบบแนะนำ page อื่นๆ ที่เกี่ยวข้องมาให้ด้วย หรือ การสร้างโมเดลเพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคต เช่น ทำนายยอดขายในไตรมาสถัดไป หรือ การทำนายว่าพนักงานคนไหนที่จะลาออกจากบริษัทในช่วง 3 เดือนข้างหน้า ตัวอย่างเหล่านี้ล้วนเป็นผลมาจากการวิเคราะห์ข้อมูลทางด้าน Data Mining

การวิเคราะห์ข้อมูลด้วย Data Mining นี้กำลังเป็นที่นิยมไปทั่วโลกด้วยแรงขับเคลื่อนอย่างหนึ่งคือ การมีซอฟต์แวร์ที่ช่วยให้ทำการวิเคราะห์ได้ง่ายขึ้น แต่ซอฟต์แวร์ส่วนใหญ่จะเป็นซอฟต์แวร์เชิงพาณิชย์ ทว่าการลงทุนซื้อซอฟต์แวร์เชิงธุรกิจเหล่านี้มาใช้งานอาจจะไม่คุ้มค่าในการลงทุนสำหรับผู้ประกอบการวิสาหกิจขนาดกลางและขนาดย่อม (SMEs) หรือ อาจารย์ นักวิจัย ดังนั้นวิธีการหนึ่งที่จะทำให้เราสามารถวิเคราะห์ข้อมูลเหล่านี้ได้คือการใช้ open source software ที่สามารถดาวน์โหลดมาใช้งานได้โดยไม่เสียค่าใช้จ่าย ซอฟต์แวร์ RapidMiner Studio 7 เป็นซอฟต์แวร์ทาง Data Mining ที่ได้รับการโหวตว่ามีผู้ใช้งานมากที่สุด ในหลักสูตรนี้จะแนะนำการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ตั้งแต่ระดับเริ่มต้น (basic) จนถึงระดับกลาง (intermediate) ด้วยการใช้ซอฟต์แวร์ RapidMiner Studio 7 ซึ่งเป็นเวอร์ชันล่าสุด

หลักสูตรนี้จะทำให้สามารถวิเคราะห์ข้อมูลด้วย Data Mining เพื่อเข้าใจพฤติกรรมผู้บริโภคของลูกค้า ทำนายอนาคตได้ สามารถทำงานวิจัยด้าน text mining หรือ image processing ได้ สามารถวิเคราะห์ข้อมูลที่ซับซ้อนได้แม้ว่าจะไม่มีความรู้เรื่องการเขียนโปรแกรมใดๆมาก่อน

หลักสูตรนี้เหมาะสำหรับ

- ผู้ที่สนใจการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining และ open source software RapidMiner Studio 7
- อาจารย์มหาวิทยาลัยที่ต้องการ update ความรู้ทางด้าน Data Mining และเรียนรู้การใช้งานซอฟต์แวร์ใหม่ๆ ที่ดีขึ้น
- บุคลากรทางด้าน IT ที่ต้องการเพิ่มทักษะทาง Data Mining
- หน่วยงานหรือองค์กรที่ต้องการนำเทคนิค Data Mining และ open source software RapidMiner Studio 7 ไปใช้ในการวิเคราะห์ข้อมูลต่างๆ ภายในองค์กรหรือทดแทน commercial software
- ผู้ประกอบการวิสาหกิจขนาดกลางและย่อม (SME) ที่ต้องการวิเคราะห์ข้อมูลลูกค้าหรือการซื้อขายสินค้า

เนื้อหาการอบรม (Course Outline)

วันที่ 1

- แนะนำการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining และการใช้ประโยชน์ในงานวิจัย
- แนะนำกระบวนการ CRISP-DM เบื้องต้นสำหรับการวิเคราะห์ข้อมูล
- แนะนำส่วนต่างๆ ของซอฟต์แวร์ RapidMiner Studio 7
- การนำเข้าข้อมูลไฟล์ Excel, CSV เข้ามาใช้ใน RapidMiner Studio 7
- ลักษณะของแอตทริบิวต์ (attribute) ต่างๆ ในชุดข้อมูล
- การเขียนไฟล์ให้อยู่ในรูปแบบของ Excel และ CSV
- การแสดงข้อมูลในกราฟแบบต่างๆ เช่น scatter plot, time series
- การค้นหา Outlier ซึ่งเป็นข้อมูลที่แตกต่างจากข้อมูลอื่นๆ
- การค้นหาข้อมูลที่ผิดพลาด (missing value) และแทนที่ด้วยค่าที่กำหนดเองหรือค่าทางสถิติ
- การแปลงข้อมูลด้วยเทคนิค discretization แบบกำหนดช่วงเองหรือแบบอัตโนมัติ
- การลดจำนวนข้อมูลด้วยการ sampling แบบต่างๆ
- การเลือกแอตทริบิวต์เพื่อใช้ในการวิเคราะห์ข้อมูล
- แนะนำการหาความสัมพันธ์ (association rules) และการประยุกต์ใช้งานด้านต่างๆ
- แนะนำเทคนิคการหาความสัมพันธ์ด้วยเทคนิค Apriori และ FP Growth
- การแปลงข้อมูลจากฐานข้อมูล relation database ให้เป็นฐานข้อมูล transaction database

- การหาความสัมพันธ์ด้วยเทคนิค FP Growth ซึ่งเป็นวิธีที่มีประสิทธิภาพมากที่สุด
- Workshop การหาความสัมพันธ์จากข้อมูลการซื้อขายสินค้าจำนวนมากกว่า 100,000 transactions ด้วย RapidMiner Studio 7
- แนะนำการแบ่งกลุ่มข้อมูล (clustering) และการประยุกต์ใช้งานด้านต่างๆ
- แนะนำตัววัดประสิทธิภาพของการแบ่งกลุ่มข้อมูล
- แนะนำการแบ่งกลุ่มข้อมูลด้วยเทคนิค K-Means และ DBScan
- Workshop การแบ่งกลุ่มข้อมูลทางการศึกษาและการแพทย์ด้วย RapidMiner Studio 7

วันที่ 2

- แนะนำการจำแนกประเภทข้อมูล (classification)
- การวัดประสิทธิภาพของการจำแนกประเภทข้อมูล
- แนะนำเทคนิค Linear Regression และการประยุกต์ใช้งาน
- การใช้งาน Linear Regression ใน RapidMiner Studio 7
- แนะนำเทคนิค Naive Bayes และการประยุกต์ใช้งาน
- การใช้งาน Naive Bayes ใน RapidMiner Studio 7
- แนะนำเทคนิค Decision Tree และการประยุกต์ใช้งาน
- การใช้งาน Decision Tree ใน RapidMiner Studio 7
- แนะนำเทคนิค K-Nearest Neighbours (KNN) และการประยุกต์ใช้งาน
- การใช้งาน KNN ใน RapidMiner Studio 7
- แนะนำเทคนิค Neural Networks และการประยุกต์ใช้งาน
- การใช้งาน Neural Networks ใน RapidMiner Studio 7
- แนะนำเทคนิค Support Vector Machines (SVM) และการประยุกต์ใช้งาน
- การใช้งาน SVM ใน RapidMiner Studio 7
- Workshop การจำแนกประเภทข้อมูลในงานด้านต่างๆ
- ด้านธุรกิจ
- ด้านการศึกษา
- ด้านการแพทย์
- การคัดเลือกแอตทริบิวต์ (attribute selection) และการประยุกต์ใช้ในการจำแนกประเภทข้อมูล
- Workshop การคัดเลือกแอตทริบิวต์และการจำแนกประเภทข้อมูลในงานด้านต่างๆ
- การจำแนกประเภทข้อมูลแบบหลายลาเบล (multi-label)
- Workshop การจำแนกประเภทข้อมูลแบบหลายลาเบล

วันที่ 3

- การเปรียบเทียบประสิทธิภาพของโมเดลจากเทคนิค classification ต่างๆ และ t-test
- แนะนำเทคนิคการหาวิธี classification ที่เหมาะสมสำหรับข้อมูลแต่ละชุด (meta-learning)
- แนะนำการทำ Text Mining ด้วย RapidMiner Studio 7
- Workshop การจำแนกข้อความที่เป็น spam จาก SMS
- Workshop การแบ่งกลุ่มข้อมูลจากข้อความรีวิว (Review)
- Workshop การหาความสัมพันธ์จากข้อความรีวิว
- แนะนำการทำ Image Mining ด้วย RapidMiner Studio 7
- Workshop การจำแนกรูปภาพออกเป็นประเภทต่างๆ