



บันทึกข้อความ

ส่วนงาน คณะวิทยาศาสตร์ หลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ โทร. ๓๘๙๐

ที่ ศธ ๐๕๒๓.๔.๖/๑๒๑

วันที่ ๑๔ กรกฎาคม ๒๕๖๐

เรื่อง ขอส่งรายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าร่วมฝึกอบรม เรื่อง “การวิเคราะห์ข้อมูลด้านเทคนิค Predictive Modeling เพื่อสนับสนุนการวิจัย”

เรียน คณบดีคณะวิทยาศาสตร์

ตามหนังสือที่ ศธ ๐๕๒๓.๔.๖/๕๘ ลงวันที่ ๑๔ มิถุนายน ๒๕๖๐ ข้าพเจ้าได้รับอนุญาตให้เข้าฝึกอบรมฝึกอบรม เรื่อง “การวิเคราะห์ ข้อมูลด้านเทคนิค Predictive Modeling เพื่อสนับสนุนการวิจัย” ระหว่างวันที่ ๒๓-๒๕ มิถุนายน ๒๕๖๐ ณ ห้องปฏิบัติการคอมพิวเตอร์ INC๒๑ สำนักดิจิทัลเพื่อการศึกษา มหาวิทยาลัยราชภัฏเชียงใหม่ นั้น

ในการนี้ ข้าพเจ้า ได้เข้าร่วมฝึกอบรมเป็นที่เรียบร้อยแล้ว จึงใคร่ขอส่งรายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าร่วมฝึกอบรมเรื่องดังกล่าว ดังรายละเอียดรายงานสรุปเนื้อหาแนบท้ายมาพร้อมกับหนังสือนี้

จึงเรียนมาเพื่อโปรดทราบ

(อาจารย์ภาณุวัฒน์ เมษะ)

ประธานอาจารย์ประจำหลักสูตรวิทยาศาสตร์บัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์

รายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าอบรม สัมมนา หรือประชุมวิชาการ

ข้าพเจ้า นายภานุวัฒน์ เมฆะ ตำแหน่งอาจารย์ สังกัด คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้ ขอ
นำเสนอรายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าร่วมอบรม เรื่อง “การวิเคราะห์ข้อมูลด้าน
เทคนิค Predictive Modeling เพื่อสนับสนุนการวิจัย” ในระหว่างวันที่ 23-25 มิถุนายน 2560
ณ ห้องปฏิบัติการคอมพิวเตอร์ INC21 สำนักดิจิทัลเพื่อการศึกษา มหาวิทยาลัยราชภัฏ เชียงใหม่ ตามหนังสือ
ขออนุญาตเดินทางไปราชการ เลขที่ ศ.ธ.0523.4.6/58 ลงวันที่ 15 มิถุนายน 2560 ซึ่งการเข้าร่วมอบรม
ดังกล่าว ข้าพเจ้าไม่ได้ใช้งบประมาณการพัฒนามูลค่า

ดังนั้นจึงขอนำเสนอรายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าร่วมอบรม ดังนี้

การวิเคราะห์ข้อมูลเทคนิค Predictive Modeling เพื่อสนับสนุนการวิจัยเชิงประยุกต์โดยซอฟต์แวร์
RapidMiner Studio 7 และ R (ขั้นปานกลางและสูง)

ซึ่งเทคนิคการจำแนกประเภทข้อมูลหรือที่เรียกว่า Predictive Modeling เป็นเทคนิคที่นิยมใช้กัน
มากในการวิเคราะห์ข้อมูลและการทำงานวิจัยเชิงประยุกต์ ซึ่งกระบวนการจำแนกประเภทข้อมูลจะแบ่งเป็น
สองส่วนคือ (1) การนำข้อมูลสอน (training data) มาสร้างโมเดลและวัดประสิทธิภาพของโมเดล และ (2)
การนำโมเดลที่ได้ไปใช้ทำนาย (predict) เพื่อหาคำตอบให้กับข้อมูลใหม่ โดยการสร้าง โมเดลนั้นมีหลายๆ
เทคนิค เช่น

1. เทคนิค Decision Tree ซึ่งสร้างโมเดลในรูปแบบของ Tree เพื่อช่วยตัดสินใจ
2. เทคนิค Naïve Bayes ซึ่งสร้างโมเดลโดยใช้การคำนวณความน่าจะเป็นจากโอกาสที่เกิดขึ้นมาแล้ว
ใน training data
3. เทคนิค K Nearest Neighbours ซึ่งสร้างโมเดลโดยการเปรียบเทียบความคล้ายคลึงกับข้อมูล
training data
4. เทคนิค Neural Network เป็นการสร้างโมเดลที่ใช้สมการคณิตศาสตร์ที่ซับซ้อนในลักษณะที่คล้าย
กับการทำงานของสมองมนุษย์
5. เทคนิค Support Vector Machines มีข้อดีโดยการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถแบ่ง
ข้อมูลด้วยโมเดลเส้นตรงได้

ทว่าในหลายๆครั้งการประยุกต์ใช้เทคนิค Predictive Modeling แบบทั่วไปอาจจะไม่ได้ให้ผลลัพธ์ที่ดี
เนื่องจากข้อมูลจริงมีความท้าทายในหลายๆ เรื่อง เช่น

- ข้อมูลที่ใช้ในการเรียนรู้มีจำนวนข้อมูลในแต่ละคลาสคำตอบแตกต่างกันเป็นอย่างมาก หรือ
เรียกว่าเป็น Imbalanced data เช่น ข้อมูลของลูกค้าในธนาคารที่มีการฉ้อโกง (fraud) จะมีจำนวนน้อยมาก

เมื่อเทียบกับลูกค่าปกติ แต่สิ่งที่เราต้องการหา คือ การทำนายว่าการทำงานของลูกค่าคนใดบ้างที่เกิดการฉ้อโกงขึ้น หรือ ข้อมูลการตอบรับโปรโมชั่นต่างๆ

- ข้อมูลมีจำนวนแอตทริบิวต์ที่เยอะและบางครั้งมีความซ้ำซ้อนและไม่จำเป็นอยู่ เช่น การสกัดข้อความต่างๆ เพื่อนำมาสร้างโมเดลจะมีค่าแตกต่างกันมาก แต่บางค่าอาจจะเกิดไม่บ่อยนักทำให้ข้อมูลส่วนใหญ่มีค่าความถี่ของค่าเป็น 0

ในหลักสูตรนี้จะเน้นการปรับปรุงประสิทธิภาพของการสร้างโมเดลการจำแนกประเภทข้อมูลเพื่อให้มีประสิทธิภาพขึ้นโดยใช้ซอฟต์แวร์ RapidMiner Studio 7 ที่เรียนรู้ได้ง่ายและเพิ่มประสิทธิภาพโดยการเชื่อมต่อกับภาษา R อีกด้วย โดยในหลักสูตรนี้ผู้เข้าร่วมอบรมจะได้เรียนรู้

- หลักการสร้างโมเดลเพื่อจำแนกประเภทข้อมูลแบบพื้นฐานต่างๆ และการวัดประสิทธิภาพของโมเดล

- การเขียนโปรแกรมภาษา R เบื้องต้นเพื่อการแสดงกราฟและการสร้างโมเดลต่างๆ

- การจัดการข้อมูลที่เป็นลักษณะ Imbalance โดยการ sampling แบบต่างๆ

- การจัดการข้อมูลที่มีแอตทริบิวต์ที่เยอะเกินความจำเป็นและซ้ำซ้อนกัน ด้วยวิธีการ Attribute Selection แบบต่างๆ

- การเพิ่มประสิทธิภาพของโมเดลด้วยการใช้หลายๆ เทคนิคร่วมกันทำงาน ด้วยวิธี Ensemble แบบต่างๆ

- การค้นหาพารามิเตอร์ (parameter) ของแต่ละเทคนิคที่เหมาะสมด้วยวิธีการ optimization