

เลขที่.....	
แบบรายงานการเข้าอบรม/ สัมมนา/ ศึกษาดูงาน ☐ รายบุคคล ☒ กลุ่มบุคคล	
ชื่อ - สกุล : นายทรงยศ ชันบุตรศรี ตำแหน่ง : นักวิชาการโสตทัศนศึกษา	
กลุ่มงาน / หน่วยงาน : ฝ่ายเทคโนโลยีห้องสมุด สำนักหอสมุดกลาง	
ชื่อ - สกุล : ธนวัฒน์ เสริฐสุวรรณกุล ตำแหน่ง : นักวิชาการคอมพิวเตอร์	
กลุ่มงาน / หน่วยงาน : ฝ่ายเทคโนโลยีห้องสมุด สำนักหอสมุดกลาง	
ชื่อหลักสูตร	โครงการอบรมเชิงปฏิบัติการเรื่อง Data Analytics with Pentaho BI, Weka, R, and Hadoop: From Business Intelligence to Data Science โมดูลที่2 Diagnostic, Predictive, and Prescriptive Analytics
วัน เดือน ปี	21-22 พฤศจิกายน 2561
สถานที่จัด	ณ ห้อง Orchid 4 ชั้น 11 โรงแรม Jasmine สุขุมวิท 23 กรุงเทพฯ
หน่วยงานผู้จัด	สมาคมศูนย์วิชาการไทย-ออสเตรเลีย
ค่าใช้จ่าย	☒ ไม่มี ☐ มี จำนวน.....บาท เบิกจ่ายจากงบประมาณ ☐ แผ่นดิน ☐ เงินรายได้ ☐ งบอื่นๆ (ระบุ)
ใบเกียรติบัตร/ วุฒิบัตร	☐ ได้รับ ☒ ไม่ได้รับ เนื่องจาก ผู้จัดอบรมได้ให้เข้าร่วมเพิ่มโดยไม่เสียค่าใช้จ่าย ☐ ไม่มี
สรุปสาระสำคัญ โครงการอบรมเชิงปฏิบัติการเรื่อง Data Analytics with Pentaho BI, Weka, R, and Hadoop: From Business Intelligence to Data Science โมดูลที่2 Diagnostic, Predictive, and Prescriptive Analytics หลักการและเหตุผล ปัจจุบันองค์กรทางธุรกิจทั้งภาครัฐบาลและเอกชนให้ความสนใจเกี่ยวกับเรื่อง Data Science, Big Data มีการเพิ่มจำนวนมากขึ้นเพราะเรื่องเหล่านี้สามารถนำมาใช้ประโยชน์ในการจัดเก็บข้อมูลและการใช้งานข้อมูลที่มีจำนวนมาก เมื่อมีการจัดโครงสร้างการเก็บข้อมูลจำนวนมากแล้ว ขั้นตอนถัดมาคือการวิเคราะห์ข้อมูลหรือ Data Analytics เป็นกระบวนการในการวิเคราะห์ข้อมูลที่มีอยู่ทั้งหมดโดยแบ่งออกเป็นหลายๆ เรื่องในการวิเคราะห์และเป็นการพัฒนาระบบ BI หรือระบบ Business Intelligence ซึ่งเป็นการวิเคราะห์ข้อมูลทางธุรกิจขององค์กรแบบกึ่งอัตโนมัติเพื่อใช้ในการสนับสนุนการบริหารและการตัดสินใจ ซึ่งข้อมูลส่วนใหญ่มาจากฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ที่เกิดจากธุรกรรม (transaction) ขององค์กร ที่เรียกว่า Operational Database เป็นข้อมูลเชิงโครงสร้าง โดยมีการออกแบบ Data model สำหรับสร้างฐานข้อมูลที่พร้อมวิเคราะห์ซึ่งอยู่ในฐานข้อมูลเดียวกันที่เรียกกันว่าคลังข้อมูล Data warehouse หรือคลังข้อมูลย่อย Data mart โดยมี Business Intelligence Platform (software) เช่น Pentaho Business Analytic Suite เป็นระบบจัดการที่ครบวงจรตั้งแต่ Data Engineering, Data Analysis & Modeling, Data Visualization ดังนั้นในการอบรมนี้เพื่อการเรียนรู้อุปกรณ์เพื่อนำมาใช้สร้างฐานข้อมูล, จัดเก็บข้อมูลที่รูปแบบคลังข้อมูล, การจัดการข้อมูลก่อนจัดเก็บลงฐานข้อมูล และการใช้งานข้อมูลในฐานข้อมูล ในขั้นพื้นฐานเพื่อนำไปประยุกต์ใช้กับแต่ละองค์กร	

วัตถุประสงค์

- เพื่อศึกษาและการพัฒนาระบบ BI จากกรณีตัวอย่าง การทำเหมืองข้อมูลของระบบการจัดการเรียนออนไลน์ ATutor
- เพื่อศึกษาและใช้งานโปรแกรม Weka และ deploy model ด้วย Pentaho Weka Scoring Plugin Data Science Pack ของ Pentaho Suite

หัวข้อการบรรยาย Data Analytics with Pentaho BI, Weka, R, and Hadoop: From Business Intelligence to Data Science

เรื่องการทำเหมืองข้อมูลของระบบการจัดการเรียนออนไลน์ : กรณีศึกษาของระบบ ATutor

โดย ผศ.ดร. สุณี รักษาเกียรติศักดิ์

- กรอบแนวคิดเชิงทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล
 - การทำเหมืองข้อมูล เป็นกระบวนการสร้างความรู้จากฐานข้อมูลประมวลผลรายการ เพื่อใช้ในการบริหาร การตัดสินใจ และการวางแผนเชิงกลยุทธ์ โดยใช้วิธีทางสถิติ หรือวิธีการเรียนรู้ของเครื่องมือ (Machine learning) โดยการพัฒนาอัลกอริทึมที่สร้างตัวแบบการเรียนรู้หรือโมเดลจากข้อมูลจำนวนมาก และประเมินตัวแบบว่ามีความถูกต้องเพียงใดจากข้อมูลอีกชุดหนึ่ง
- ประเภทของโจทย์ปัญหา ที่ทำเหมืองข้อมูล
 - การสรุปข้อมูล (descriptive)
 - เป็นการสรุปข้อมูลให้ได้สารสนเทศที่น่าไปใช้ได้ โดยนำเสนอด้วยสถิติเช่น จำนวนร้อยละ, ค่ากลางกับค่าการกระจาย ค่าเปอร์เซ็นต์ไทล์ต่างๆ อาจจะแสดงผลในรูปแบบของกราฟเป็นต้น
 - การจำแนกข้อมูล (classification)
 - เป็นการสร้างโมเดลจำแนกข้อมูลจากคุณสมบัติของข้อมูล โดยต้องรู้ประเภทของข้อมูล ผลลัพธ์ที่ได้จะเป็นกฎเพื่อระบุประเภทของข้อมูล ตัวอย่างอัลกอริทึม เช่น ID3, J48 เป็นต้น
 - การพยากรณ์ (prediction)
 - เป็นการสร้างโมเดลสำหรับพยากรณ์คุณสมบัติของข้อมูลใดข้อมูลหนึ่ง ตัวอย่างอัลกอริทึม เช่น Multiple Linear Regression ความรู้ที่ได้จะเป็นสมการพยากรณ์
 - การแบ่งกลุ่มข้อมูล (clustering)
 - เป็นการแบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่มโดยใช้คุณสมบัติของข้อมูลเป็นตัวแบ่ง ตัวอย่างอัลกอริทึม เช่น Simple K-Mean ผลลัพธ์ที่ได้จะเป็นการจัดกลุ่มข้อมูลได้ centroid ของกลุ่มเป็นค่าเฉลี่ยของคุณสมบัติต่างๆ
 - การหาความสัมพันธ์ของข้อมูล (Association)
 - เป็นการหาความสัมพันธ์ของคุณสมบัติของข้อมูลว่า มีคุณสมบัติข้อใดที่ตรงกัน ตัวอย่างต้นแบบคือ โจทย์การวิเคราะห์ตะกร้าสินค้า เพื่อพิจารณาสินค้าที่ผู้ซื้อพร้อมกัน ตัวอย่างอัลกอริทึม เช่น Apriori ผลลัพธ์ที่ได้จากกฎสัมพันธ์แบบ IF กับ THEN
- มาตรการกระบวนการทำเหมืองข้อมูล มี 6 ขั้นตอน
 - Business Understanding คือการเข้าใจโจทย์ที่ต้องการการทำเหมืองข้อมูล
 - Data Understanding คือการรู้เรื่องของข้อมูลที่สามารถนำมาตอบโจทย์ รวมถึงแหล่งที่มาข้อมูล

- Data preparation คือการจัดทำข้อมูลที่ต้องการนำมาวิเคราะห์ ต้องมีการแปลงข้อมูลก่อนนำมาใช้
 - Modeling คือการสร้างโมเดลเพื่อนำมาวิเคราะห์ข้อมูล
 - Evaluation คือการประเมินโมเดลว่าสามารถนำมาใช้วิเคราะห์ได้ผิดพลาดมากหรือน้อยได้
 - Deployment คือการนำผลลัพธ์ไปใช้ในด้านต่างๆ เช่น การพยากรณ์ เป็นต้น
- วัตถุประสงค์ของการวิจัย
- เพื่อประยุกต์ใช้หลักการทำเหมืองข้อมูล (data mining) สร้างความรู้ (knowledge) จากฐานข้อมูลประมวลผลรายการ (transaction) ของระบบจัดการการเรียนแบบออนไลน์ (Learning Management System หรือ LMS) ATutor ของมหาวิทยาลัยศรีนครินทรวิโรฒ
 - เพื่อให้ได้สารสนเทศที่ผู้บริหารมหาวิทยาลัยสามารถใช้อำหนดแผนกลยุทธ์ที่จะส่งเสริม กระตุ้น สร้างแรงจูงใจให้หน่วยงานและคณาจารย์ใช้การเรียนการสอนแบบออนไลน์ผสมผสานกับการเรียนการสอนแบบปกติให้เกิดเป็นการจัดการเรียนการสอนแบบไฮบริด (Hybrid) มากขึ้น
- โจทย์สำหรับทำ BI ของระบบ e-learning
- ข้อ 1: จำนวนวิชาใน ATutor ของแต่ละคณะ (Faculty) ในแต่ละภาคการศึกษา (Semester) แยกตามประเภทวิชา (CourseType) ว่าเป็นวิชาตามหลักสูตรของมหาวิทยาลัย หรือวิชาตามอัธยาศัย
 - ผลการวิเคราะห์ข้อมูลเดิมจะเป็นการนำเสนอ Descriptive Statistics (Cross tab) ดูผลในตารางที่ 2 (ใช้โปรแกรม R วิเคราะห์)
 - ใน BI จะนำเสนอแบบ OLAP สามารถดูผลลัพธ์ได้แบบ interactive online
 - ข้อ 2: มีการใช้งานในลักษณะต่าง ๆ มากน้อยเพียงใด เช่น การสร้างเพจเนื้อหา การสร้างข้อสอบออนไลน์ การสร้างประกาศ การสร้างแหล่งเรียนรู้รายวิชา การมอบหมายงาน/การบ้าน ผลการเข้าใช้งานของนิสิต เช่น จำนวนนิสิตที่ลงทะเบียนเรียน จำนวนครั้งของการเข้าใช้วิชา
 - ผลการวิเคราะห์ข้อมูลจะเป็นการนำเสนอ Descriptive Statistics
 - ดูผลในตารางที่ 2 (ใช้โปรแกรม R วิเคราะห์)
 - ใน BI สามารถใช้ Pentaho Data Integration
 - ข้อ 3: ในรายวิชาตามหลักสูตรของมหาวิทยาลัยทั้งหมด (SWU Course) สามารถจัดกลุ่มวิชาเพื่อบอกสถานะของการใช้งานว่ามีการใช้งานมากน้อยเพียงใด เช่น เป็นวิชาที่ไม่ค่อยได้ใช้งาน วิชาที่มีการใช้งานในระดับปานกลาง หรือ วิชาที่มีการใช้งานในระดับมาก เพื่อสามารถใช้เป็นเกณฑ์ในการสร้างแรงจูงใจในการให้ภาระงาน
 - ทำ Clustering ของวิชา SWU Course โดยใช้ Simple K-means (ใช้ Weka)
 - Problem type: Clustering
 - 5 clusters: InAct, ActL1, ActL2, ActL3, ActL4, ActL5
 - ผลลัพธ์แสดงใน Problem type: Clustering
 - ผลการวิเคราะห์ข้อมูลเดิมจะเป็นการนำเสนอ Descriptive Statistics (Cross tab) ดูผลในใช้โปรแกรม R วิเคราะห์
 - ใน BI จะนำเสนอแบบ OLAP สามารถดูผลลัพธ์ได้แบบ interactive online (SWUCourse Cube)

- ข้อ 4: จำนวนอาจารย์ที่ใช้ ATutor ของแต่ละคณะ (Faculty) แยกตามสถานะการใช้งาน (Status) เกณฑ์: ใน 1 ปี ถ้ามีวิชาที่ Active อย่างน้อย 1 วิชา อาจารย์นั้นจะได้ Status เป็น Active
 - ผลการวิเคราะห์ข้อมูลเดิมจะเป็นการนำเสนอ Descriptive Statistics (Cross tab) ใช้โปรแกรม R วิเคราะห์
 - ใน BI จะนำเสนอแบบ OLAP สามารถดูผลลัพธ์ได้แบบ interactive online (Instructor Cube)
 - หมายเหตุ มีการปรับเกณฑ์ใหม่: ใน Semester หนึ่ง ๆ ถ้ามีวิชาที่ Active อย่างน้อย 1 วิชา อาจารย์นั้นจะได้ Status เป็น Active
- ข้อ 5: จากข้อ 3 สร้างโมเดลของการจัดกลุ่มว่าการที่จะจำแนกวิชาว่ามีการใช้งานในระดับใด ใช้เกณฑ์ใดมาจับ
 - Problem type: Classification
- ข้อ 6: แอดทริบิวต์หรือตัวแปรใดที่ส่งผลต่อการเข้าใช้งานของวิชาบ้าง
 - Problem type: Prediction
- ฐานข้อมูลที่ใช้
 - ฐานข้อมูลของระบบ ATutor ในหนึ่งปีการศึกษา ตั้งแต่วันที่ 2007-10-25 ถึงวันที่ 2008-10-17 ซึ่งมีข้อมูลของสองภาคการศึกษาคือ ภาคการศึกษาที่ 2/2550 และภาคการศึกษาที่ 1/2551
- สรุป โจทย์สำหรับทำ Cube
 - Course: จำนวนวิชา (CourseCount) และจำนวนนักเรียนที่ลงทะเบียนเรียน (Enroll) ของแต่ละคณะ (Faculty) ในแต่ละภาคการศึกษา (Semester) แยกตามประเภทวิชา (CourseType) ว่าเป็นวิชาตามหลักสูตรของมหาวิทยาลัย หรือวิชาตามอัธยาศัย
 - SWU Course: บอกสถานะของการใช้งานในรายวิชาตามหลักสูตรของมหาวิทยาลัยว่าแยกตามสถานะการใช้งาน (Status) แยกตามคณะ (Faculty)
 - Instructor: จำนวนอาจารย์ที่ใช้ ATutor ของแต่ละคณะ (Faculty) แยกตามสถานะการใช้งาน (Status)
- การเตรียมข้อมูลสำหรับการวิเคราะห์
 - เป็นกระบวนการที่ยากที่สุด และใช้เวลามากที่สุด อาจจะต้องวนลูปการทดสอบหลายครั้งมาก
 - เรียกว่าเป็นกระบวนการ ETL (Extract Transform and Load)
 - เมื่อได้วิธีการที่ต้องการและถูกต้องแล้ว จะได้ job สำหรับ execute และ Load เข้า Data warehouse สำหรับ data ชุดใหม่ได้อัตโนมัติ

ประโยชน์ที่ได้รับ

- กรอบแนวคิดเชิงทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล
- ได้รู้และเข้าใจความหมาย, ความสำคัญ ประโยชน์ของการทำเหมืองข้อมูลเพื่อใช้ในการบริหาร การตัดสินใจ และการวางแผนเชิงกลยุทธ์ ของแต่ละองค์กร

- ได้รู้และเข้าใจการบางประเภทของโจทย์ปัญหา ที่ทำเหมืองข้อมูล
 - การสรุปข้อมูล (descriptive)
 - เป็นการสรุปข้อมูลให้ได้สารสนเทศที่น่าไปใช้ได้ โดยนำเสนอด้วยสถิติเช่น จำนวนร้อยละ, ค่ากลางกับค่าการกระจาย ค่าเปอร์เซ็นต์ไทล์ต่างๆ และแสดงผลในรูปแบบของกราฟ
 - การจำแนกข้อมูล (classification)
 - เป็นการสร้างโมเดลจำแนกข้อมูลจากคุณสมบัติของข้อมูล โดยต้องรู้ประเภทของข้อมูล ผลลัพธ์ที่ได้จะเป็นกฎเพื่อระบุประเภทของข้อมูล
 - การพยากรณ์ (prediction)
 - เป็นการสร้างโมเดลสำหรับพยากรณ์คุณสมบัติของข้อมูลใดข้อมูลหนึ่ง
 - การแบ่งกลุ่มข้อมูล (clustering)
 - เป็นการแบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่มโดยใช้คุณสมบัติของข้อมูลเป็นตัวแบ่ง
 - การหาความสัมพันธ์ของข้อมูล (Association)
 - เป็นการหาความสัมพันธ์ของคุณสมบัติของข้อมูลว่า มีคุณสมบัติข้อใดที่ตรงกัน
- มาตรการกระบวนการทำเหมืองข้อมูล มี 6 ขั้นตอน
 - Business Understanding ได้รู้และเข้าใจการตั้งโจทย์ในการทำเหมืองข้อมูล
 - Data Understanding ได้รู้และเข้าใจการนำข้อมูลที่สามารถนำมาตอบโจทย์
 - Data preparation ได้รู้และเข้าใจการจัดทำข้อมูลที่ต้องการนำมาวิเคราะห์
 - Modeling ได้รู้และเข้าใจการสร้างโมเดลเพื่อนำมาวิเคราะห์ข้อมูล
 - Evaluation ได้รู้และเข้าใจวิธีการประเมินโมเดลว่าสามารถนำมาใช้วิเคราะห์ได้หรือไม่
 - Deployment ได้รู้และเข้าใจการนำผลลัพธ์ไปใช้ในด้านต่างๆ เช่น การพยากรณ์ เป็นต้น
- การใช้งานโปรแกรมในการวิเคราะห์ข้อมูล
 - Pentaho data integration เป็นเครื่องมือในการจัดการข้อมูลในฐานข้อมูลให้เชื่อมโยงกันได้ หรือเรียกว่า ETL
 - Pentaho schema workbench ใช้ในการสร้าง OLAP, Data cube เพื่อใช้งานด้านสนับสนุนการวิเคราะห์ข้อมูลที่อยู่ในคลังข้อมูล ในหลายรูปแบบ เช่น ตารางข้อมูล, Dashboard ,รายงาน เป็นต้น
 - โปรแกรม R ใช้ในการสร้างโมเดลเพื่อการวิเคราะห์ข้อมูล, OLAP, Data Cube

นำความรู้ที่ได้รับมาใช้ปรับปรุงการทำงาน ดังต่อไปนี้	
หัวข้อการปรับปรุง / พัฒนา	รายงานผลการปรับปรุง/ พัฒนา ภายในวันที่

ข้อเสนอแนะอื่นๆ(ถ้ามี)

ข้อเสนอแนะอื่นๆ(ถ้ามี)

ผู้รายงาน

(นายธนวัฒน์ เสรีสุวรรณกุล)

ตำแหน่ง นักวิชาการคอมพิวเตอร์

วันที่ 19 ธันวาคม 2561

ผู้รายงาน

(นายทรงยศ ชันบุตรศรี)

ตำแหน่ง นักวิชาการโสตทัศนศึกษา

วันที่ 19 ธันวาคม 2561

ความคิดเห็นของผู้บริหารที่กำกับดูแลหน่วยงาน

นางอัมพร ใจผ่องแผ้ว รองผู้อำนวยการ
โรงเรียนโสตทัศนศึกษา

ลงชื่อ.....

(อาจารย์สาโรช เมลาพันธ์)

วันที่.....

21 ธ.ค. 2561

- หมายเหตุ : 1. จัดทำรายงานฯ หลังจากเข้าอบรม / สัมมนา / ศึกษาดูงาน ภายใน 15 วัน เสนอหัวหน้าหน่วยงาน
2. หัวหน้าหน่วยงานเสนอความเห็น ภายใน 3 วัน เสนอผู้บริหารที่กำกับดูแล
3. แจ้งผู้รายงานทราบ และจัดเก็บเข้าแฟ้ม รายงานการเข้าอบรม / สัมมนา / ศึกษาดูงาน
4. หัวหน้าหน่วยงานติดตามผลการปรับปรุงพัฒนา