

สรุปความรู้จากโครงการฝึกอบรมหลักสูตร
หลักสูตร “การใช้เครื่องมือดิจิทัลเพื่อการทำงานภาครัฐ (Essential Digital Tools for Workplace)”

ผู้สอน รองศาสตราจารย์ ดร.วราภรณ์ ร่มสายหยุด
สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

โดย นางสาวกรวิกา รัตนพนันท์
นักวิชาการเกษตรชำนาญการพิเศษ
กลุ่มวิชาการเพื่อการพัฒนาที่ดิน

วัตถุประสงค์

1. เพื่อให้ผู้เรียนมีความรู้พื้นฐานเกี่ยวกับข้อมูลขนาดใหญ่ (Big Data)
2. เพื่อให้ผู้เรียนมีความรู้พื้นฐานเกี่ยวกับเครื่องมือวิเคราะห์ข้อมูล (Hadoop) เพื่อการทำงานบริหารข้อมูลขนาดใหญ่
3. เพื่อให้ผู้เรียนมีความเข้าใจพื้นฐานเกี่ยวกับการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อการบริหารภาครัฐ

เป็นการเรียนรู้เกี่ยวกับความหมายของ Big Data และนำไปใช้ให้เกิดประโยชน์ กระบวนการในการจัดเก็บ Big Data รวมถึงเรียนรู้เกี่ยวกับเครื่องมือที่ใช้ในการวิเคราะห์ Big Data และแนวทางในการวิเคราะห์ Big Data เพื่อการทำงานภาครัฐ

เนื้อหาประกอบด้วย

Big Data คือ ข้อมูลที่มีขนาดใหญ่อยู่ในรูปแบบข้อมูลที่มีขนาดใหญ่อยู่ในรูปแบบที่หลากหลายและมีการเปลี่ยนแปลงอย่างรวดเร็ว รูปแบบของข้อมูล Big Data มี ๕ รูปแบบ ได้แก่ Behavioral Data, Image and sounds, Languages, Records, และ Sensors

ข้อมูล Big Data ประกอบไปด้วยคุณลักษณะ ๔ ประการ คือ

๑. Volume (มีปริมาณและขนาดใหญ่)
๒. Variety (มีความหลากหลายสามารถเป็นได้ทั้งข้อมูลที่มีโครงสร้างและข้อมูลที่ไม่สามารถจับ pattern ได้ไม่ว่าจะเป็นข้อมูลที่เป็นตัวหนังสือ, รูปภาพ, ข้อมูลเสียงที่ถูกบันทึกไว้, วิดีโอหรือไฟล์ประเภทอื่นจากหลากหลายแหล่งที่มา)
๓. Velocity (คือข้อมูลที่มีการเปลี่ยนแปลงอัปเดตตลอดเวลาอย่างรวดเร็ว มีการส่งผ่านข้อมูลอย่างต่อเนื่องในลักษณะ streaming แบบ Real-Time ทำให้การวิเคราะห์ข้อมูลแบบ manual มีข้อจำกัด)
๔. Veracity (คุณภาพของข้อมูล คือคุณภาพของข้อมูลที่สามารถนำไป วิเคราะห์ข้อมูลต่อไปได้อีกเป็นข้อมูลที่ยังไม่ผ่านการประมวลผลอยู่ในรูปแบบของข้อมูลดิบซึ่งสามารถ นำไปประมวลผลต่อไปได้และเป็นข้อมูลที่มาจกหลากหลายแหล่งไม่ว่าจะเป็น Facebook, Youtube, Twitter ซึ่งข้อมูล จากแหล่งเหล่านี้ยากที่จะสามารถควบคุมคุณภาพรวมถึงการคัดกรองข้อมูล และความน่าเชื่อถือของข้อมูล จึงต้องนำข้อมูลเหล่านี้เข้าสู่กระบวนการทำ Data Cleansing

กระบวนการทำงานของ Big Data ขั้นตอนและกระบวนการทำงานของ Big Data มีอยู่ขั้นตอนหลักๆ ดังนี้

๑. จัดเก็บข้อมูล (Storage)

การเก็บข้อมูลเป็นเรื่องสำคัญมากเรื่องหนึ่งของคนที่ข้อมูลอยู่ในมือเหมือนสสารที่ต้องการที่อยู่และหลายๆ บริษัทก็พยายามจะใช้ข้อมูลขนาดใหญ่ (Big Data) ที่เกิดขึ้นให้เป็นประโยชน์ได้มากที่สุดโดยทั่วไป การเก็บข้อมูลจะอยู่ในรูปแบบฐานข้อมูลแบบเดิม (Traditional Database) เป็นแบบที่มีการเก็บข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) และมีขนาดไม่ใหญ่มาก ซึ่งมีความเร็วในการประมวลผลไม่มากนัก สำหรับการเก็บข้อมูลขนาดใหญ่ในปัจจุบัน เรามักจะใช้ Data Warehouse หรือ Data Lake เป็นที่เก็บข้อมูล ทั้งสองอันนี้มักจะถูกพูดถึงและเปรียบเทียบกันอยู่เสมอ ทั้งที่จริงๆ แล้วมีความแตกต่างกันมาก ทั้งด้านโครงสร้าง การประมวลผลไปจนถึงคนและเหตุผลในการใช้งาน ขั้นตอนการจัดเก็บรวบรวมข้อมูลทั้งหมดจากแหล่งต่าง ๆ ไม่ว่าจะข้อมูลที่มีคุณภาพรวมถึงข้อมูลที่คาดว่าอาจจะเป็นประโยชน์ ไม่ว่าจะเก็บข้อมูลตัวอักษร ไฟล์เอกสาร ไฟล์รูปภาพ ไฟล์วิดีโอ ไฟล์เสียงที่ถูกบันทึกจะถูกเก็บรวบรวมไว้ที่นี้

Data Lake เกิดขึ้นเนื่องจากการนำเอาข้อมูลจากแหล่งข้อมูลภายนอกองค์กรมาใช้มากขึ้น ปริมาณข้อมูลจากแหล่งภายนอกมีการเพิ่มขึ้นอย่างต่อเนื่องและเพื่อแก้ไขข้อจำกัดหลายอย่างของ Data Warehouse ที่ใช้มานานแล้วข้อมูลที่จัดเก็บใน Data Lake ข้อมูลที่มีจำนวนมากขนาดใหญ่ ข้อมูลที่ไม่มีรูปแบบที่แน่นอน การเข้าถึงข้อมูลไม่สามารถเข้าถึงได้ง่าย จำเป็นต้องใช้เจ้าหน้าที่วิเคราะห์ข้อมูล ที่มีความรู้ความสามารถเฉพาะเข้ามาจัดเก็บข้อมูล (Data Scientist)

๒. การประมวลผลข้อมูล (Processing)

การประมวลผลข้อมูลหลังจากที่นำข้อมูลมารวบรวมไว้ได้ในที่เดียวแล้วข้อมูลต่างๆ จะถูกนำไปจัดหมวดหมู่ให้อยู่ในกลุ่มที่มีความเกี่ยวข้องกันหรือความสัมพันธ์ใกล้เคียงกัน ให้ผลลัพธ์คล้ายคลึงกันมากที่สุด แล้วจึงนำมาเปลี่ยนเป็นรูปแบบข้อมูลเพื่อนำเอาข้อมูลที่มีอยู่เหล่านี้เข้าระบบข้อมูลที่ผ่านการประมวลผลแล้ว

๓. การวิเคราะห์ข้อมูล (Analyst)

การวิเคราะห์ข้อมูลและนำเสนอข้อมูลหลังจากที่ข้อมูลทั้งหมดได้ถูกจัดกลุ่มและแยกประเภทเรียบร้อยแล้ว นั้น ต่อจากนั้นจะนำมาวิเคราะห์หา Pattern ความเกี่ยวข้องกันทั้งหมดที่อาจมองไม่เห็นได้เลยด้วยตา ไม่ว่าจะเป็นการทำแนวโน้มของการตลาด ความต้องการของลูกค้ากระแสที่อาจเกิดขึ้นได้ในอนาคต และข้อมูลด้านอื่นที่เป็นประโยชน์ และจัดมานำเสนอในรูปแบบต่างๆ ไม่ว่าจะป็นรูปภาพหรือกราฟ

Data Analytics รูปแบบของการวิเคราะห์ข้อมูล คือการนำข้อมูลที่มีอยู่มาเข้าสู่กระบวนการวิเคราะห์เพื่อนำข้อมูลไปใช้ประโยชน์ในด้านต่าง ๆ ที่ครอบคลุมในทุกอุตสาหกรรม ทั้งภาครัฐและภาคเอกชน ซึ่งจะต้องอาศัยเทคโนโลยีหรือซอฟต์แวร์เฉพาะด้านที่เข้ามาช่วยให้การประมวลผลและวิเคราะห์ข้อมูลเป็นไปอย่างแม่นยำ และนำไปใช้งานได้มีประสิทธิภาพ โดยสามารถแบ่งประเภทของการวิเคราะห์ข้อมูลได้ ๔ รูปแบบ ตั้งแต่ Descriptive Analytics คือการวิเคราะห์ว่าเกิดอะไรขึ้น, Diagnostic Analytics รอกการวิเคราะห์ว่า สิ่งนั้น ๆ เกิดขึ้นเพราะอะไร, Predictive Analytics คือวิเคราะห์เพื่อคาดการณ์แนวโน้มของสิ่งที่จะเกิดต่อไป และ Prescriptive Analytics คือการวิเคราะห์สิ่งที่จะเกิดขึ้นจากการใช้ทางเลือกต่าง ๆ

ประโยชน์ของการใช้ Big Data

Big Data สามารถนำไปประยุกต์ใช้ได้กับหลายภาคส่วนไม่ว่าจะเป็นภาครัฐหรือภาคเอกชนในส่วนนี้จะพูดถึงประโยชน์ของ Big Data หากนำมาปรับใช้ในภาคธุรกิจแล้วจะสามารถทำให้สามารถ

- เข้าใจลูกค้าได้มากยิ่งขึ้นด้วยการนำฐานข้อมูลที่มีอยู่ใน Big Data ศึกษาถึงลักษณะพฤติกรรมของผู้บริโภคที่มีการตัดสินใจในการเลือกสินค้าอย่างไรสามารถนำเสนอสินค้าที่มีคุณมีอยู่ให้ตรงกับความต้องการของลูกค้าได้หรือไม่หากไม่มีจะสามารถนำสินค้าชนิดอื่นที่มีอยู่นำไปทดแทนได้หรือไม่

- สามารถวิเคราะห์ความต้องการของตลาดในอนาคตได้ ข้อมูลต่างๆที่ถูกค้นหาในอินเทอร์เน็ต รวมถึงใน Social Media ต่างๆ สามารถนำมาประมวลได้ ว่ามีอะไรที่เป็นกระแสหรือได้รับความนิยมอยู่ในขณะนั้นหรือกระแสที่มีการนำมาค้นหาหรือกล่าวถึงอยู่มากที่สุดก็สามารถข้อมูลนั้นมาวิเคราะห์และวางแผนก่อนหากมีแผนที่ดีและสามารถทำได้อย่างรวดเร็วก็สามารถเป็นผู้นำกระแสได้อย่างไม่ยาก จากการใช้ข้อมูลจาก Big Data

- การวางแผนในอนาคตมีประสิทธิภาพมากยิ่งขึ้น จากข้อมูลที่มีอยู่สามารถนำผลวิเคราะห์จาก Big Data เข้ามาช่วยประกอบการวางแผนและการตัดสินใจได้เป็นอย่างดี ทั้งในเรื่องของการลด ต้นทุนเพิ่มผลผลิต โดยการเก็บข้อมูลต่างๆ ภายในองค์กรเพื่อนำไปวิเคราะห์ไม่ว่าจะเป็น กระบวนการผลิต ข้อมูลการใช้วัตถุดิบ จะทำให้สามารถทราบได้ว่าปัญหาภายในองค์กรมีหรือไม่ ต้องปรับปรุงแก้ไขส่วนใดเพื่อที่จะสามารถแก้ไขปัญหาได้อย่างรวดเร็ว เพื่อป้องกันความผิดพลาดในการผลิต

- คาดการณ์ปัญหาที่อาจเกิดขึ้น จากการนำข้อมูลที่มีจาก Big Data มาคาดการณ์ความต้องการของตลาด ซึ่งนอกจากคาดการณ์ในอนาคตได้แล้วนั้น ก็ยังสามารถนำข้อมูลส่วนนั้นมาวิเคราะห์ต่อยอดได้อีกว่าในอนาคตนั้นจะมีเหตุการณ์อะไรที่สามารถเกิดขึ้นได้บ้าง ก็สามารถนำข้อมูลส่วนนั้นนำไปวางแผนปรับนโยบายวิธีการบริหารองค์กร เพื่อให้สามารถแก้ไขปัญหาได้อย่างรวดเร็ว ที่อาจเกิดขึ้นได้ในอนาคต

- ลดงบประมาณในการลงทุนด้าน IT ในอนาคตหลังจากที่นำ Big Data มาใช้ในองค์กรแล้ว สามารถลดต้นทุนการใช้งบประมาณในด้าน IT ได้เป็นอย่างดี เนื่องจากสามารถนำข้อมูลที่ได้มา ไปใช้ประโยชน์ในด้านอื่นๆ ได้อีกพร้อมยังช่วยลดต้นทุนในการจ้างพนักงานในด้าน IT ได้อีกหลายตำแหน่งที่สามารถใช้ Big Data มาทดแทนได้