

หลักสูตร Data Science for Business

จัดโดย เขตอุตสาหกรรมซอฟต์แวร์ประเทศไทย

ระหว่างวันที่ 4 - 5 กรกฎาคม 2560

วิทยากร : อาจารย์วีรศักดิ์ กฤษณะ

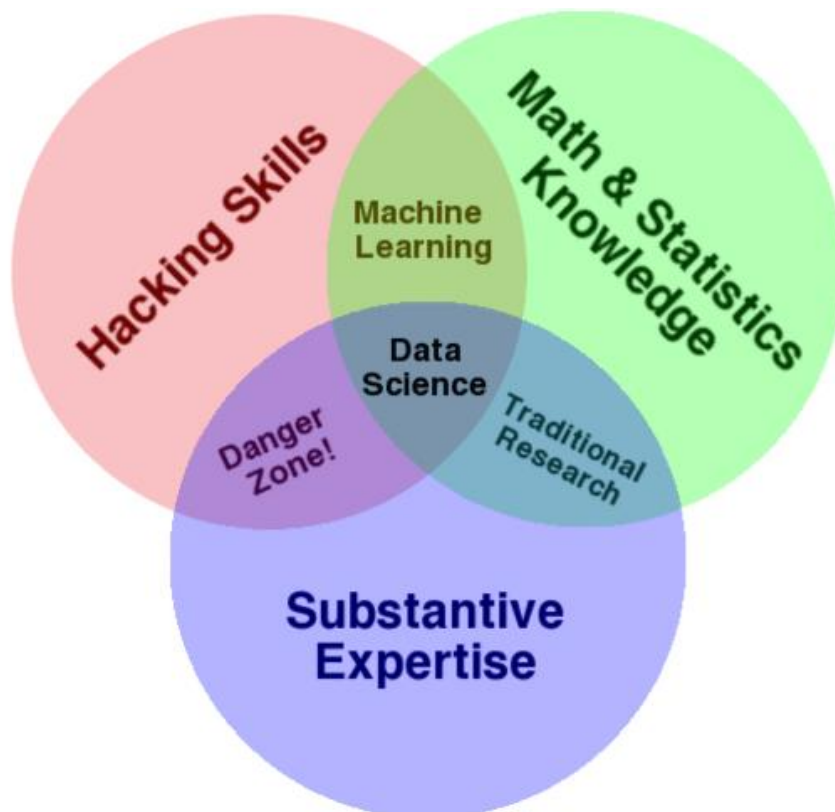
หลักสูตรนี้ออกแบบมาเพื่อเสริมสร้างความรู้ความเข้าใจเกี่ยวกับวิทยาศาสตร์ข้อมูล (Data Science) และให้ความเข้าใจถึงประโยชน์ของการนำข้อมูล (data) ไปใช้ในด้านการพยากรณ์ให้เกิดประโยชน์กับองค์กร ทั้งทางด้านกลยุทธ์องค์กรและการบริหารงาน

หลักสูตรนี้เป็นหลักสูตรเบื้องต้น ที่ปูพื้นฐานความรู้และอธิบายถึง Data Science ในแต่ละด้าน โดยกล่าวถึงประวัติศาสตร์ การนำไปใช้ประโยชน์ ตลอดจนวงจรชีวิตของ Data Science โดยผู้เรียนสามารถนำไปปรับประยุกต์ใช้กับงานของตนเองได้

วิทยาศาสตร์ข้อมูล (Data Science) หมายถึง วิธีการที่จะได้มาซึ่งความรู้จากข้อมูลขนาดใหญ่อย่างมีประสิทธิภาพ (รวดเร็วและแม่นยำ) และชาญฉลาด (ตรงประเด็นและเป็นประโยชน์) โดยการวิเคราะห์หรือค้นหา สิ่งที่ยังไม่รู้ในข้อมูลนั้น

Data Science มีองค์ประกอบที่สำคัญ 3 ส่วน ได้แก่

- 1) ความรู้ด้านคณิตศาสตร์และสถิติ (Math & Statistics Knowledge)
- 2) ทักษะในการเจาะลึกข้อมูลอย่างมีประสิทธิภาพ (Hacking Skills)
- 3) ความเชี่ยวชาญในเรื่องที่จะวิเคราะห์ (Substantive Expertise)



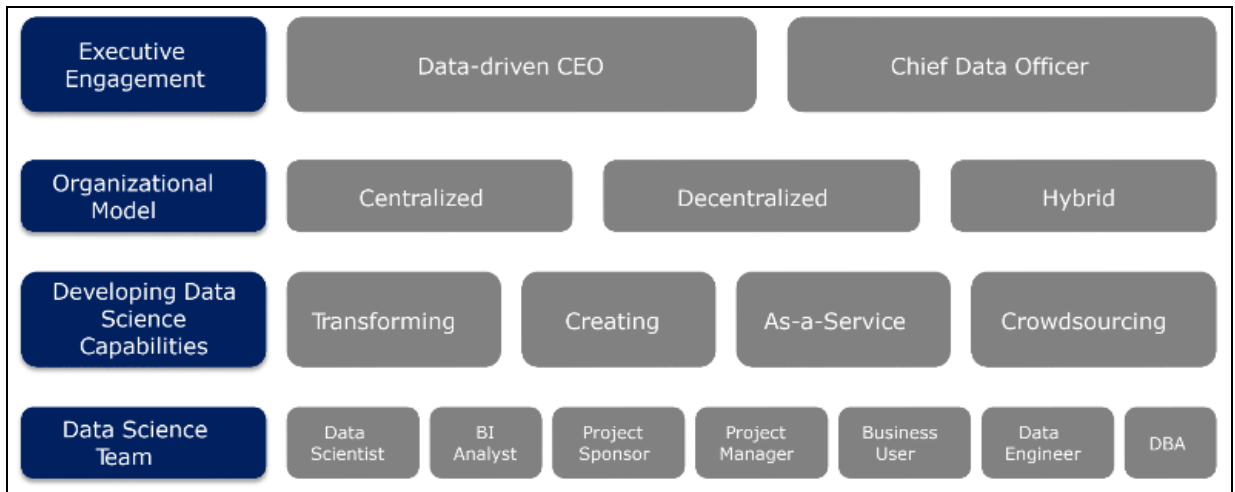
นักวิทยาศาสตร์ข้อมูล (Data Scientist) หมายถึง บุคลากรที่มีความรู้ในด้าน Data Science และสามารถใช้เทคโนโลยีทางด้าน Cloud Computing, Big Data, Machine Learning, Data Mining, Statistics และ Internet of Things เพื่อบริหารจัดการข้อมูลและสกัดความรู้จากข้อมูลขนาดใหญ่ได้

คุณลักษณะพิเศษของนักวิทยาศาสตร์ข้อมูล (Data Scientist) ต้องเป็นบุคคลที่มีความอยากรู้อยากเห็น มีความคิดริเริ่มสร้างสรรค์ ไม่ยึดติดกับมุมมองเดิมๆ มีความมุ่งมั่นในการออกแบบและทดสอบที่ทำเป็นประจำ เรียนรู้จากความล้มเหลว และใส่ใจในรายละเอียดทุกขั้นตอน

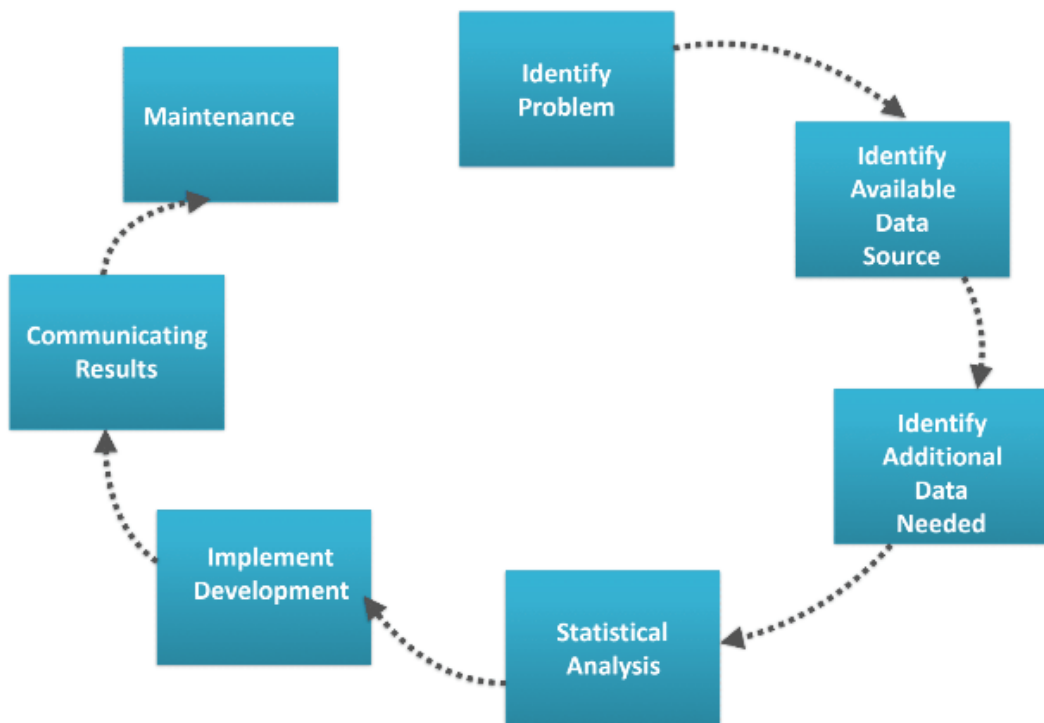
นักวิทยาศาสตร์ข้อมูลสมัยใหม่ (Modern Data Scientist) จึงต้องมีคุณสมบัติดังนี้



เนื่องจากคุณสมบัติข้างต้นนี้ไม่สามารถหาได้ในบุคคลเพียงคนเดียว ดังนั้น องค์กรจึงต้องสร้างทีมงานวิทยาศาสตร์ข้อมูล (Data Science Teams) โดยมีกรอบการพัฒนาดังนี้

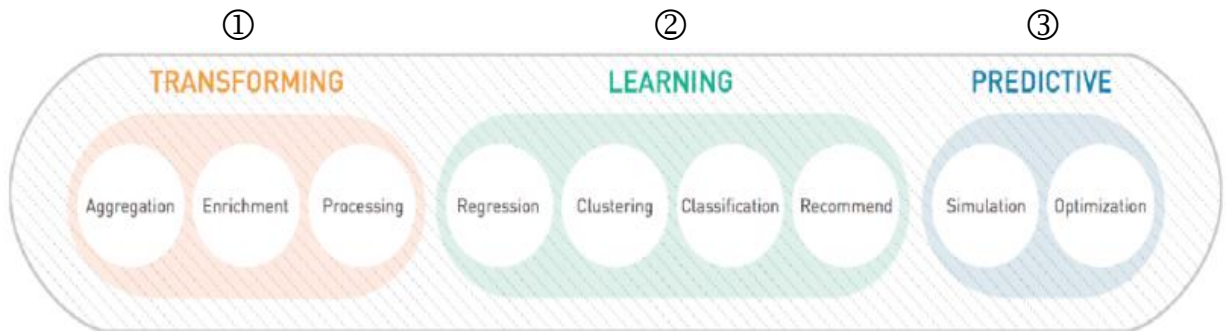


วงจรชีวิตของวิทยาศาสตร์ข้อมูล (Data Science Life Cycle)



1. ระบุปัญหา (Identify the problem) ที่ต้องการศึกษาจากข้อมูลที่มีอยู่
2. ระบุแหล่งข้อมูลที่มี (Identify available data source) หรือที่สามารถหาได้
3. ระบุความจำเป็นที่ต้องใช้ข้อมูลจากแหล่งอื่นๆ เพิ่มขึ้น (Identify if additional data sources are needed)
4. วิเคราะห์ค่าสถิติ (Statistical Analysis) จากข้อมูลที่มีอยู่
5. นำไปใช้ และพัฒนาปรับปรุง (Implementation, Development)
6. ส่งผ่านผลลัพธ์ที่ได้ (Communicate results) ไปยังผู้ใช้ข้อมูลที่เหมาะสม
7. บำรุงรักษาระบบและข้อมูล (Maintenance)

ระดับชั้นของเทคนิคในการวิเคราะห์ข้อมูล มีดังนี้



1. Transform Analytics

- **Aggregation** เทคนิคในการหาผลรวมของข้อมูล โดยใช้ค่าสถิติพื้นฐาน เช่น ค่าเฉลี่ย (mean) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) เป็นต้น
- **Enrichment** เทคนิคในการเพิ่มสารสนเทศให้กับข้อมูล เช่น แหล่งข้อมูล
- **Processing** เทคนิคในการประมวลผล เพื่อจัดเตรียมและตัดแยกข้อมูล

2. Learning Analytics

- **Regression** เทคนิคในการประมาณค่าความสัมพันธ์ระหว่างตัวแปร รวมถึงความเข้าใจว่าตัวแปรมีความสำคัญต่อการพยากรณ์ค่าในอนาคต
- **Clustering** เทคนิคการแยกกลุ่มข้อมูล
- **Classification** เทคนิคการจัดกลุ่มสมาชิก
- **Recommendation** เทคนิคการพยากรณ์อัตราหรือการอ้างอิงกับพฤติกรรมหรือข้อมูลในอดีต

3. Predictive Analytics

- **Simulation** เทคนิคการจำลองสถานการณ์ เพื่อพยากรณ์จากพฤติกรรมภายใต้เงื่อนไขใหม่
- **Optimization** เทคนิควิจัยการดำเนินงานที่เน้นการเลือกองค์ประกอบที่ดีที่สุดจากทางเลือกที่มี

ตัวอย่างของการประยุกต์ใช้ Data Science กับข้อมูลขนาดใหญ่ เช่น การแสดงผล Google Map

