

Data Preprocessing คือการจัดการข้อมูลก่อนการนำไปประมวลผล

1.Cleaning Data ทำการทำความสะอาดข้อมูล มีความจำเป็นเพราะว่าข้อมูลในโลกนี้มีความสกปรกและไม่สะอาด สาเหตุมาจากเครื่องมือไม่ดีอาจจะทำให้ข้อมูลผิดเพี้ยนได้ เช่น ข้อมูลขาดหาย หรือข้อมูลไม่สอดคล้องวันเดือนปีกัน

การจัดการกับ Missing

- ลบค่า missing ออก
- นั่งดูตัวข้อมูลและพิจารณาเติมเองหรือลบออก
- กรอก Missing ตัวใหม่ขึ้นมาเพื่อตอบค่าที่ไม่รู้

2.Data Integration การรวมข้อมูลมาจากหลายแหล่งมารวมเป็นตารางเดียวกัน

Data Preprocess ทำเพื่อแก้ไขข้อมูลส่วนที่ผิดพลาด ขาดหาย หรือทำการรวมข้อมูลจากหลายแหล่งเข้าด้วยกัน เพื่อความแม่นยำตามจุดประสงค์ที่ต้องการจะใช้