

**รายงาน**

**ปัญหา Classification และ Regression**

**จัดทำโดย**

นายศุภณัฐ บุญสารี 593020466-7

นายพรเทพ เนตรเดชา 593020930-8

**เสนอ**

ผศ.ดร.สิรภัทร เชี่ยวชาญวัฒนา

**322497 Special Topics in Computer Science ( Data Analytics and Mining)**

**ภาคเรียนที่ 1 ปีการศึกษา 2561**

**ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์**

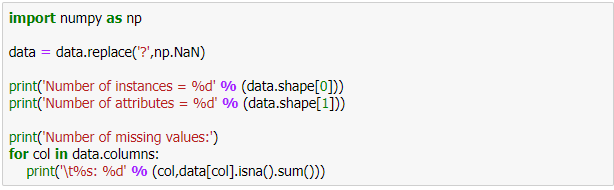
**มหาวิทยาลัยขอนแก่น**

**ปัญหา Classification**

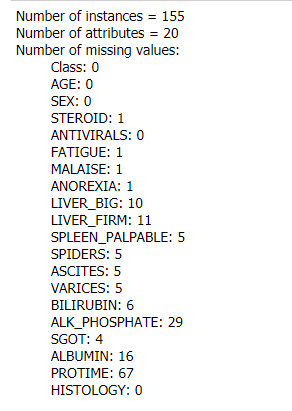
นำเข้าข้อมูลจาก UCI Machine Learning ที่ชื่อ hepatitis โดยใช้คำสั่ง



วิเคราะห์ความเสียหายของข้อมูล



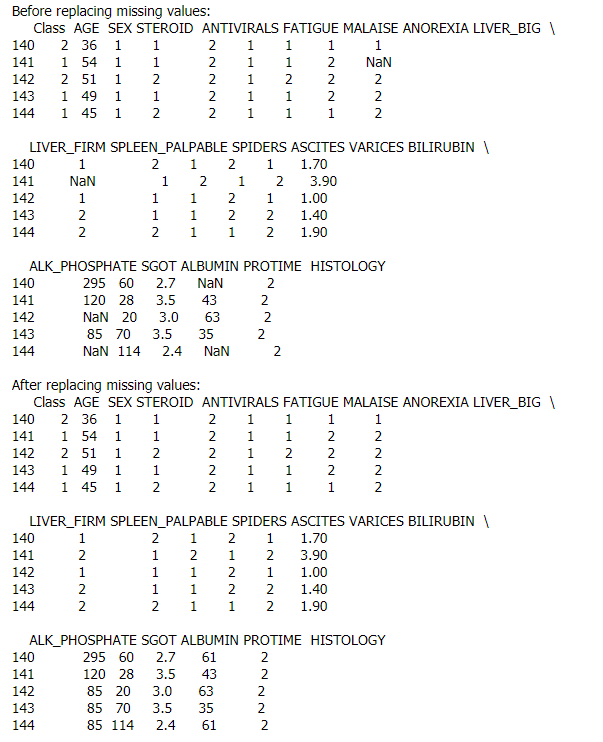
จะได้ผลลัพธ์ดังนี้ :



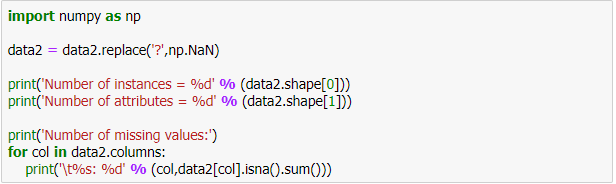
จัดการกับข้อมูลโดยการเติมข้อมูลเข้าไปแทนที่ข้อมูลที่เป็น ‘?’ โดยใช้ median เพื่อให้ค่าที่เป็น ? หายไป



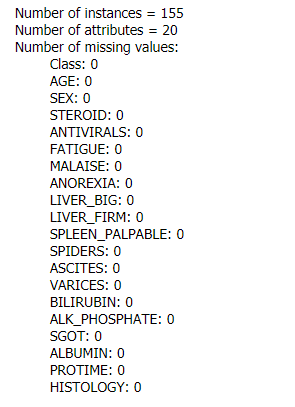
จะได้ผลลัพธ์ดังนี้ :



วิเคราะห์ข้อมูลอีกครั้งเพื่อดูว่าความเสียหายของข้อมูลได้หายไปหรือยัง



จะได้ผลลัพธ์ดังนี้ :



ค่าเป็น 0 ทั้งหมดถือว่าไม่มี missing value แล้ว

แบ่งข้อมูลเพื่อทำ Training data และ Testing data



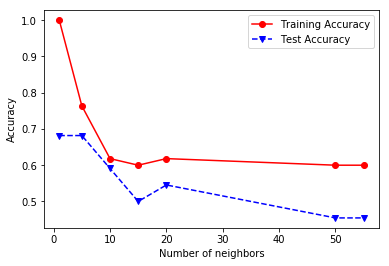
โดยใช้ข้อมูลเพื่อนำไป test = 42 ตัว

จากนั้นทำการ training และ testing data โดยใช้ KNeighborsClassifier และ linear\_model

**การทำ KNeighborsClassifier**



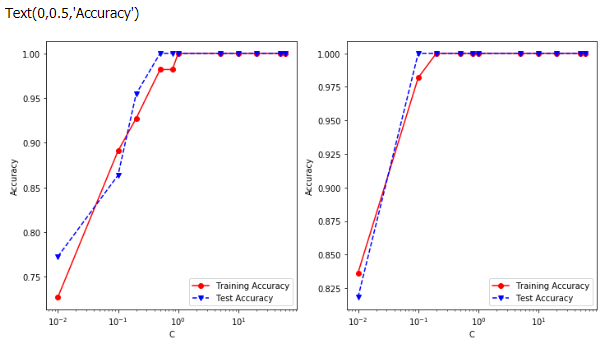
จะได้ผลลัพธ์ดังนี้ :



**การทำ linear\_model**

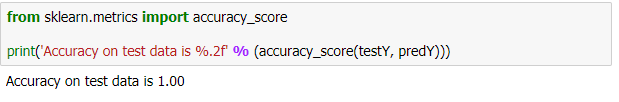


จะได้ผลลัพธ์ดังนี้ :



**สรุป**

จากภาพข้างต้นจะได้ว่า การ train ข้อมูลโดยการเติมข้อมูลเข้าไป ให้ผลที่มี accuracy ที่สูงที่สุด



**ปัญหา regression**

จากข้อมูลสภาพอากาศซึ่งทำการตรวจวัดจากสถานี ภาคตะวันออกเฉียงเหนือ จังหวัดขอนแก่น ตั้งแต่ปี 2550 – 2561 โดยนำเข้าข้อมูล 4 ชนิด ดังนี้ 1. ก๊าซซัลเฟอร์ไดออกไซด์

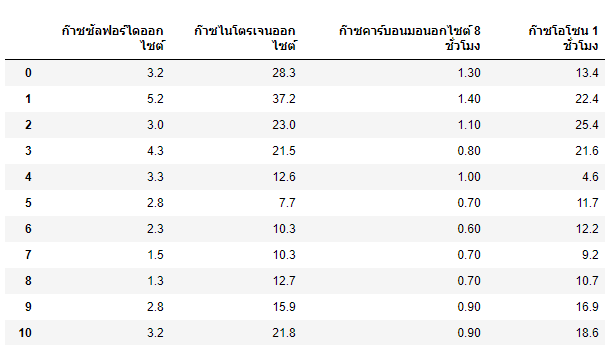
2. ก๊าซไนโตรเจนไดออกไซด์

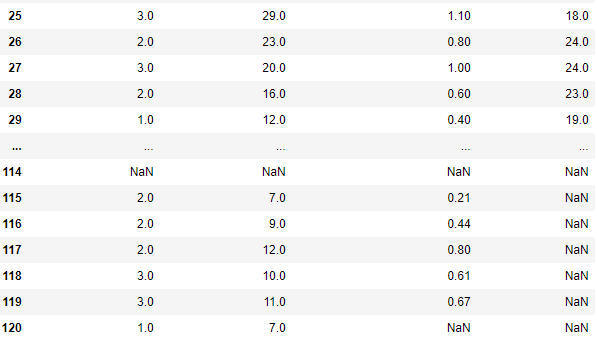
3. ก๊าซคาร์บอนมอนอกไซด์

4. ก๊าซโอโซน

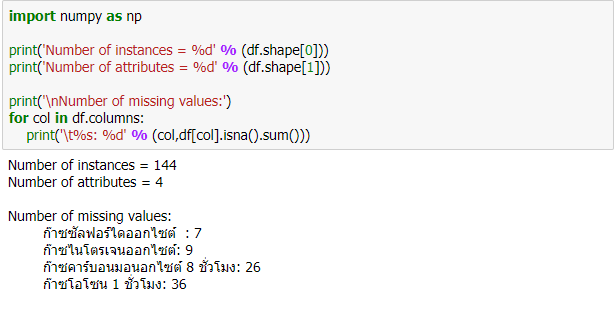


จะได้ผลลัพธ์ดังนี้ :

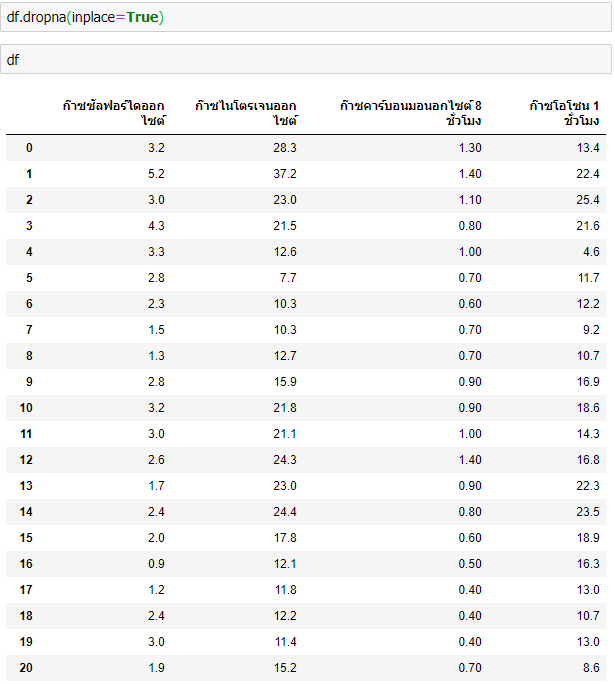




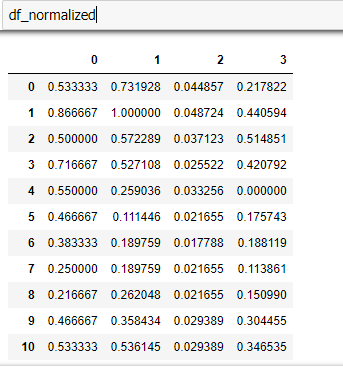
วิเคราะห์ความสูญหายของข้อมูล

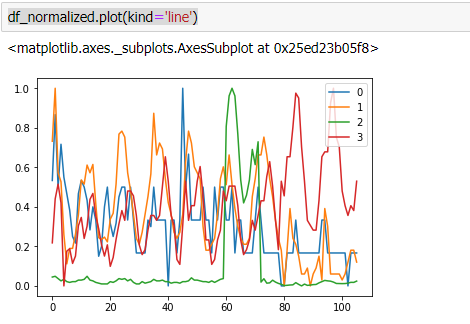


จัดการกับข้อมูลโดยการตัดทิ้งออกไป

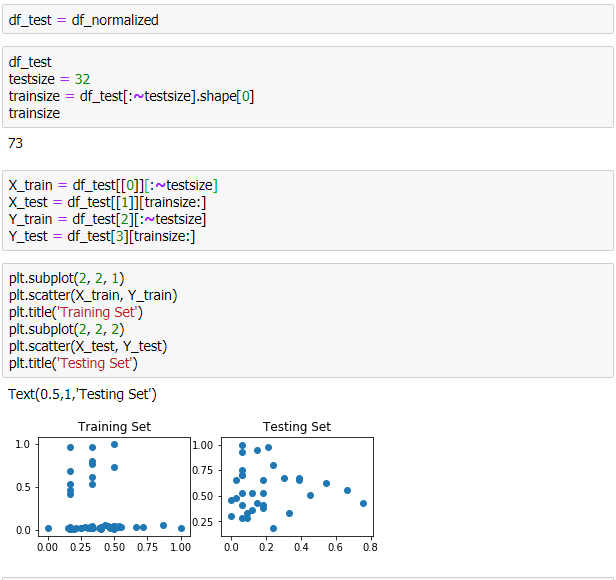


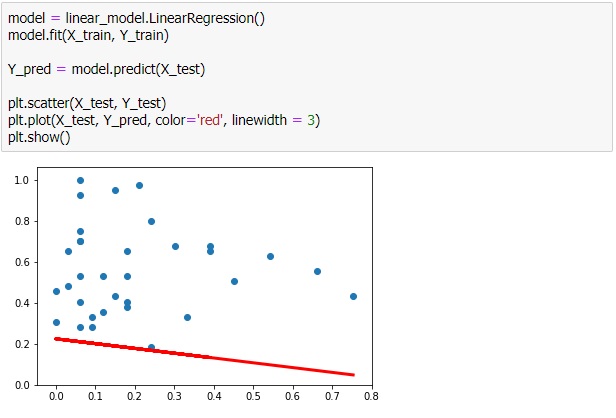
แปลงข้อมูลให้อยู่ในรูปอย่างง่ายเพื่อสามารถนำไปสร้างตัวแบบ ทำการแปลงข้อมูลโดยการใช้ normalize ให้อยู่ในช่วง 0 – 1



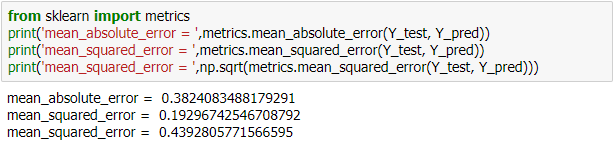


จัดแบ่งข้อมูลเพื่อนำไป training และ testing data





สร้างตัวแบบโดยใช้ค่าประสิทธิภาพ Mean Square Error(MSE)



**สรุป**

การสร้างตัวแบบโดยใช้ค่าประสิทธิภาพ Mean Sqare Error(MSE) ได้ค่า = 0.4392805771566595