

**เรื่อง**

**จัดทำโดย**

นางสาวกาณจนาภรณ์ พรมทิศ รหัสนักศึกษา 025830491052-8

นางสาวศรุตา ชื่นบุญเชิด รหัสนักศึกษา 025830491059-3

นายชนะโชค แก้ไข รหัสนักศึกษา 025830491098-1

**เสนอ**

อาจารย์สุวนีย์ กลุกรินีธรรม

รายงานเล่มนี้เป็นส่วนหนึ่งของการเรียนวิชาการทำเหมืองข้อมูล

ภาคเรียนที่ 2 ปีการศึกษา 2560

มหาวิทยาลัยเทคโนโลยีสารสนเทศราชมงคลตะวันออก วิทยาเขตจักรพงษภูวนารถ

**ขั้นตอนการทำเหมืองข้อมูล 6 ขั้นตอน**

**1.Business Understanding**

* บ้านพักคนชราแห่งหนึ่งต้องการทำนายว่าผู้ที่อาศัยในบ้านพักมีโอกาสเป็นโรคหัวใจขาดเลือด จะได้มีการดูแลเป็นพิเศษ

**2.Data Understanding**

* ข้อมูลเก่าของผู้สูงอายุที่ป่วย 138 คน
* ผู้ป่วยแต่ละรายมีจำนวนแอตทริบิวต์ทั้งหมด 8 แอตทริบิวต์
* ข้อมูลผู้สูงอายุได้ถูกรวบรวมเก็บไว้ใน trumb drive

- ไฟล์ Disdase\_testing.csv ใช้สำหรับสร้างโมเดล

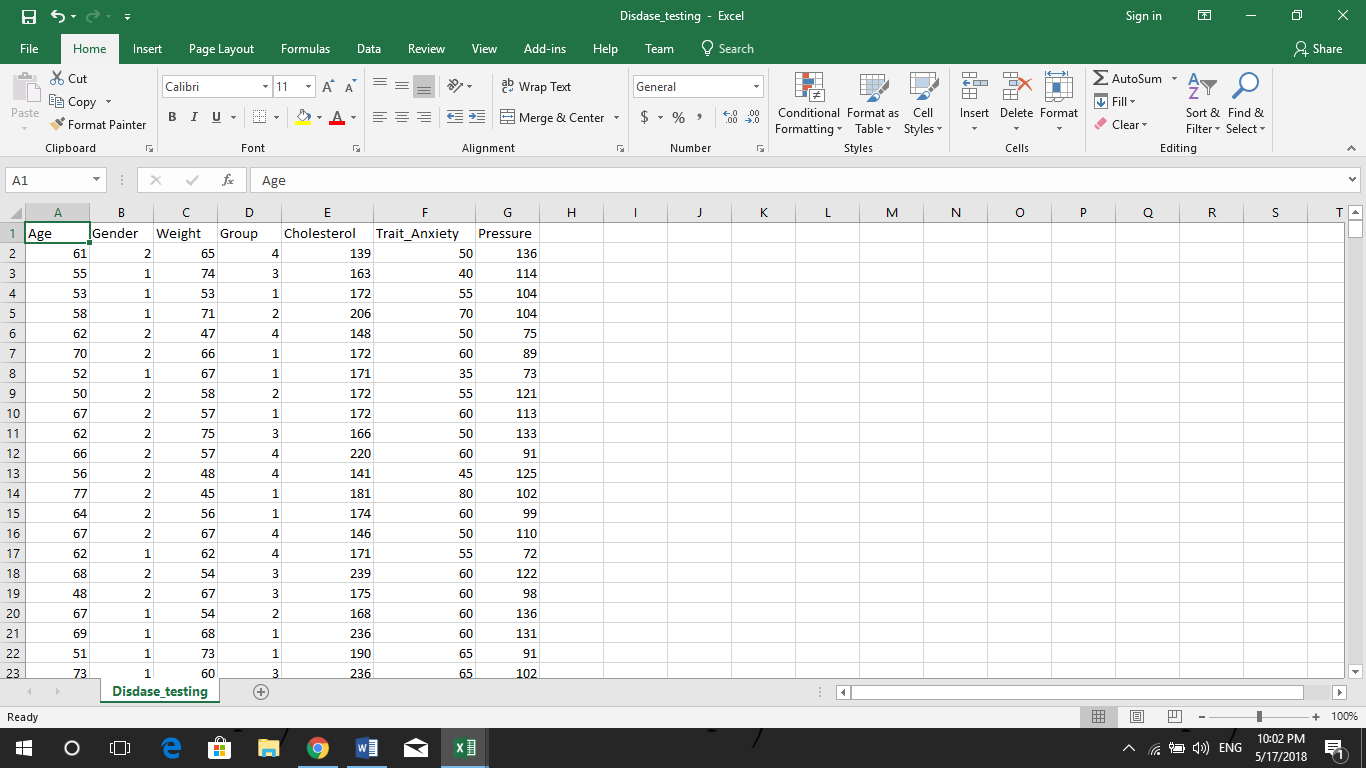
- ไฟล์ Disdase\_training.csv ใช้สำหรับสร้างมาทำนาย

- แอตทริบิวต์ต่าง ๆ ในชุดข้อมูล

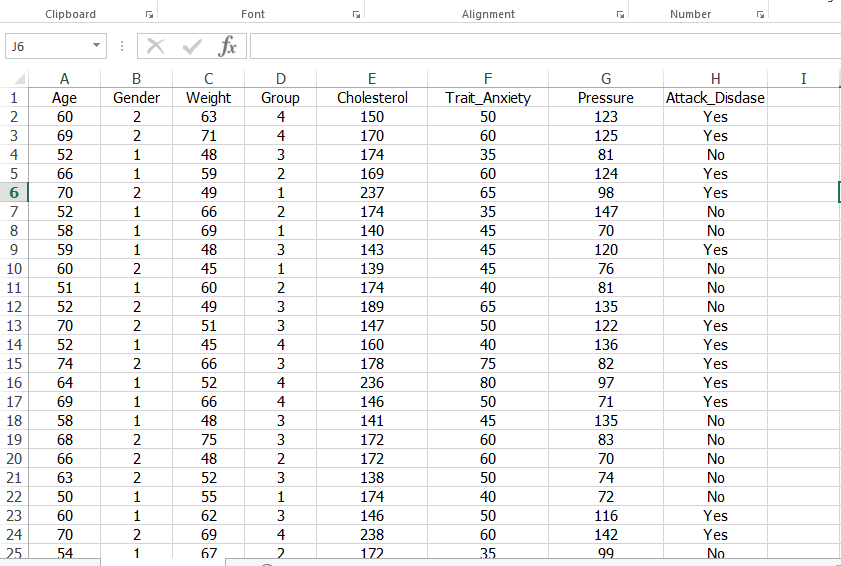
|  |  |  |
| --- | --- | --- |
| แอตทริบิวต์ | ประเภท | ความหมาย |
| Age | Interger | อายุของผู้สูงอายุ |
| Gender | Interger | เพศของผู้สูงอายุ |
| Weight | Interger | น้ำหนักของผู้สูงอายุ |
| Group | Interger | หมู่เลือดของผู้สูงอายุ |
| Cholesterol | Interger | ปริมาณคลอเลสตอรอล |
| Trait\_Anxiety | Interger | ระดับความเครียด |
| Pressure | Interger | ระดับความดันของโลหิต |
| Ischemic\_heart | Binominal | โรคหัวใจขาดเลือด |

|  |  |  |
| --- | --- | --- |
| Gender | 1 | เพศชาย |
| 2 | เพศหญิง |
| Group | 1 | กรุ๊ปA |
| 2 | กรุ๊ปAB |
| 3 | กรุ๊ปB |
| 4 | กรุ๊ปO |

**ตัวอย่างไฟล์ข้อมูล**



**ข้อมูลชุด Disdase\_testing.csv**



**ข้อมูลชุด Disdase\_training.csv**

**3. Data Preparation**

* ไม่มีข้อมูลที่ผิดพลาด

**4. Modeling**

* สร้างโมเดล K-Nearest Neighbor

**5. Evaluation**

* ทดสอบประสิทธิภาพของโมเดลด้วยวิธีการ 10-fold cross-validation

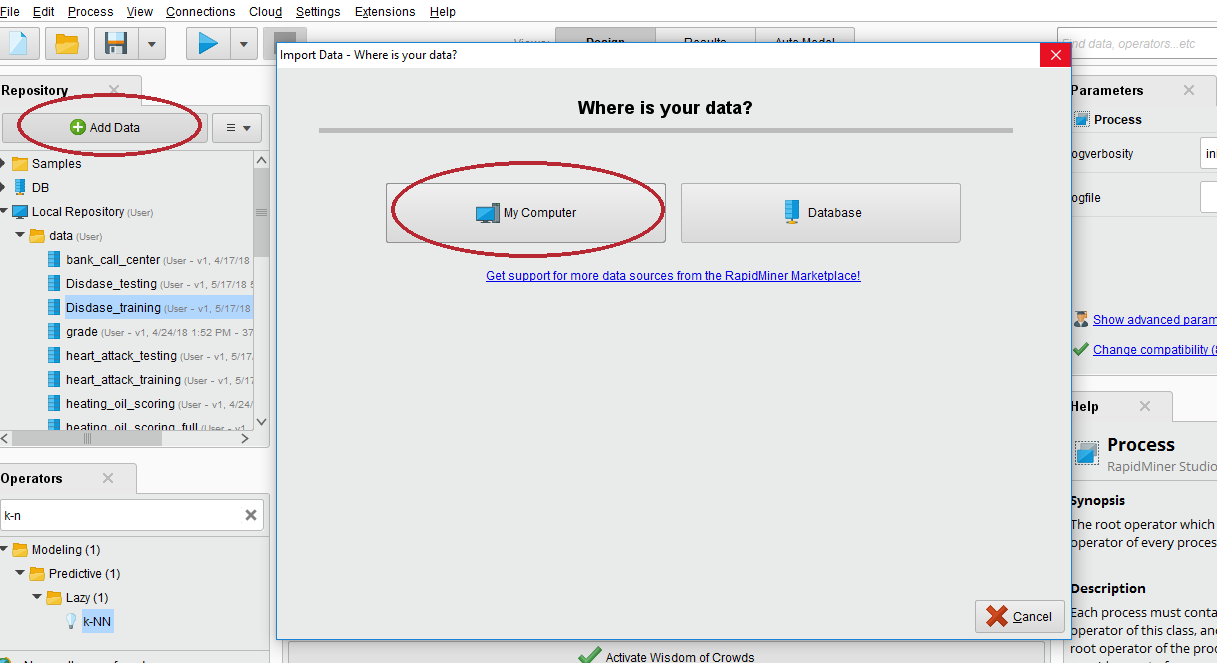
**6. Deployment**

* แนะนำการดูแลตัวเองและการดูแลรักษาได้ประสิทธิภาพที่ตรงจุด

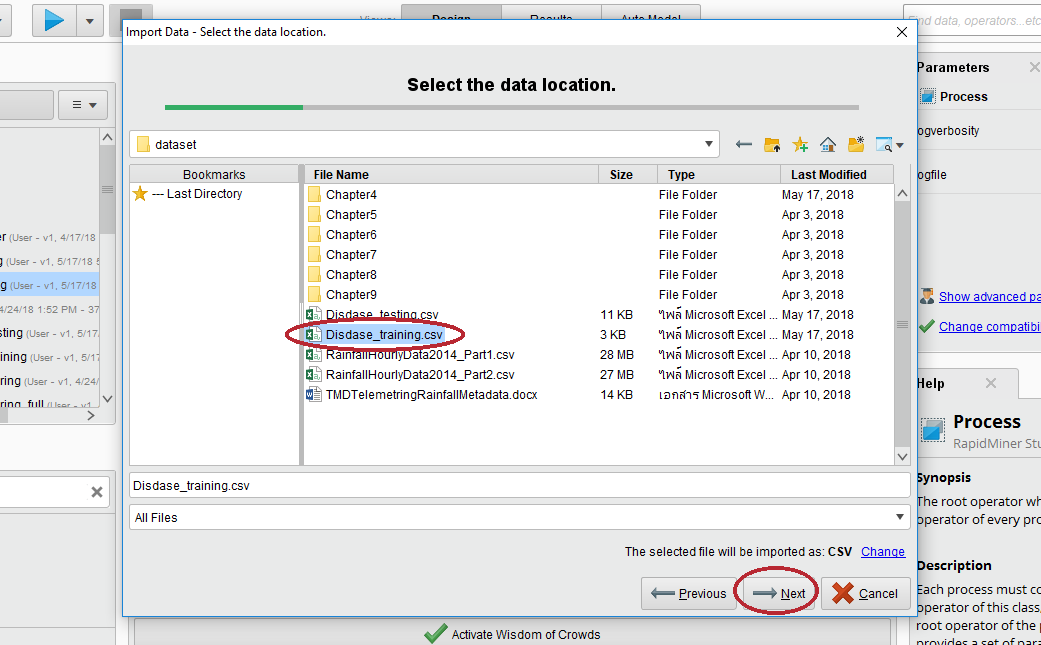
**โอเปอเรเตอร์ที่ใช้**

|  |  |  |
| --- | --- | --- |
| รูป | ชื่อ | คำอธิบาย |
|  | Read CSV | ใช้สำหรับอ่านไฟล์ CSV |
|  | x-Validation | แย่งข้อมูลสำหรับสร้างโมเดลและทดสอบโมเดล แบบ 10-fold cross-validation |
|  | K-NN | ใช้สำหรับสร้างโมเดล K-Nearest Neighbors |
|  | Apply Model | ใช้สำหรับ predict ข้อมูลใหม่ |
|  | Performance | สำหรับแสดงตัวชี้วัดของโมเดล classification |

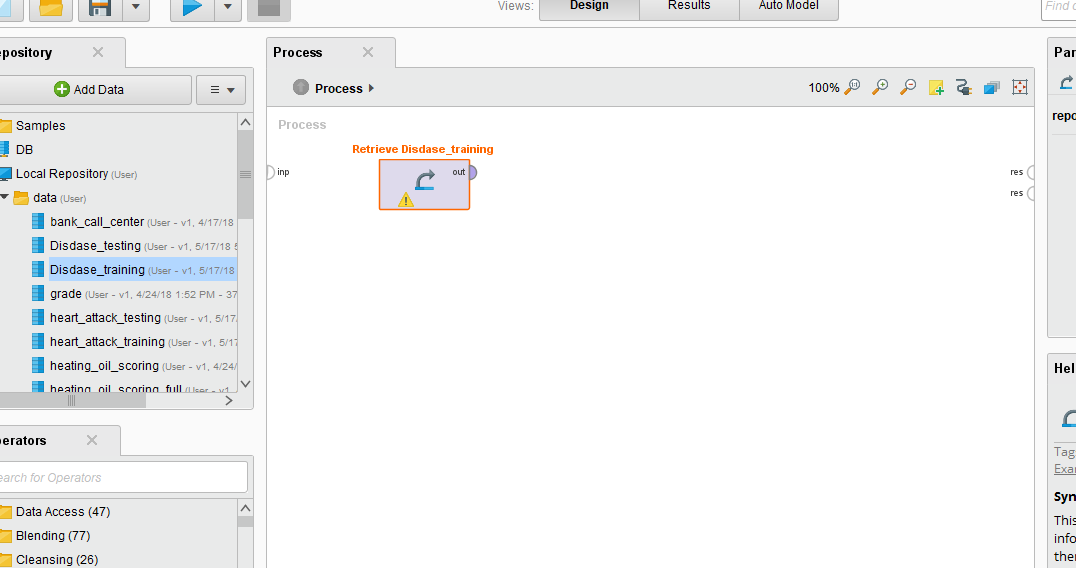
**ขั้นตอนการสร้างโมเดล**



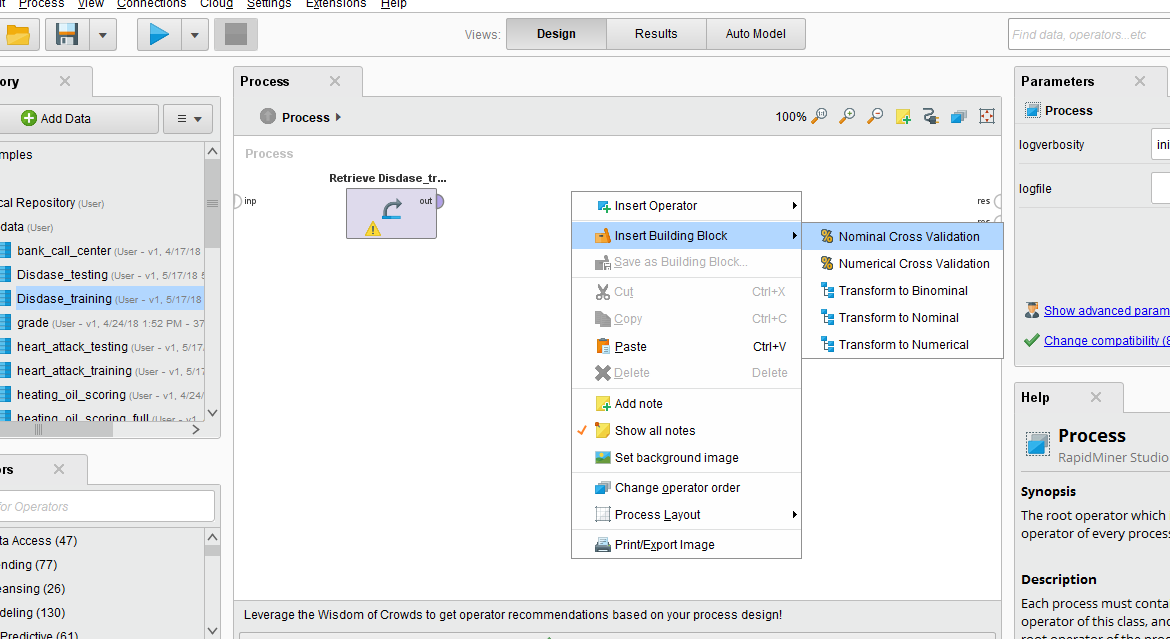
รูปที่ 1. Add dataโดยเลือกจากไฟล์ที่ใช้ในการเก็บข้อมูล



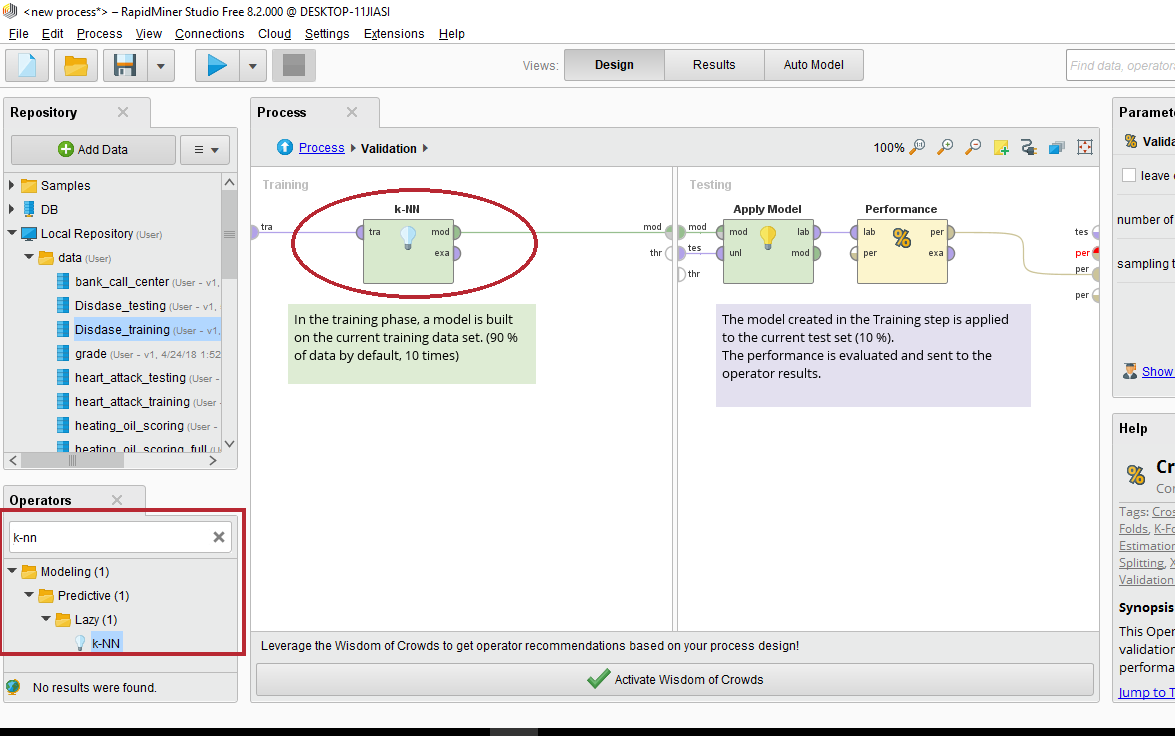
รูปที่ 2. เลือกไฟล์ Disdase\_training .csv เอามาใช้



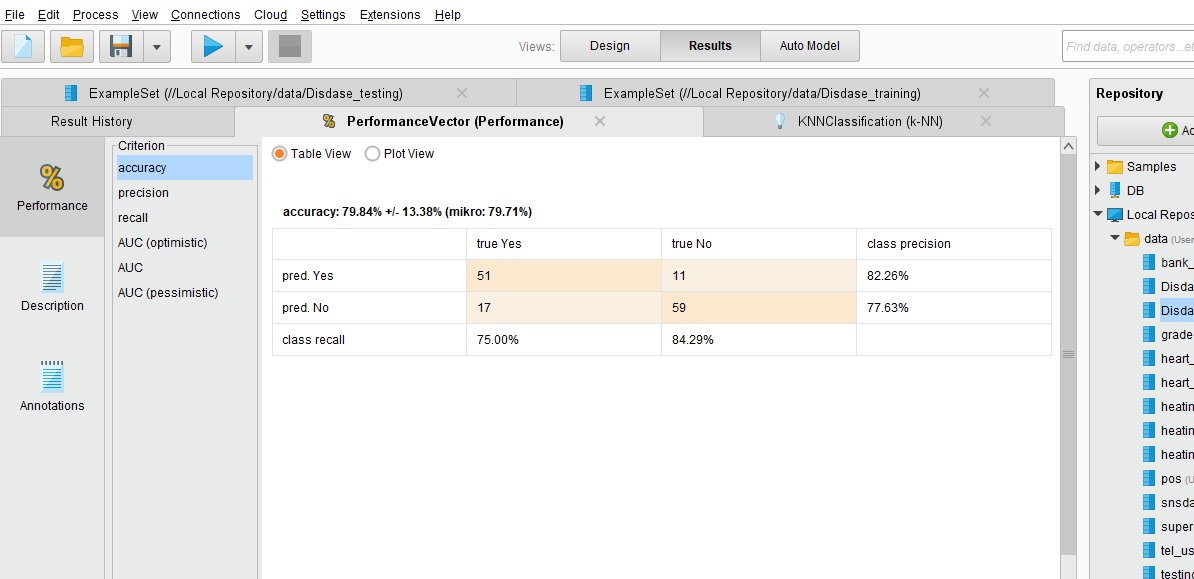
รูปที่ 3. เลือก Retrieve ใส่ไฟล์ disdase\_training.csv



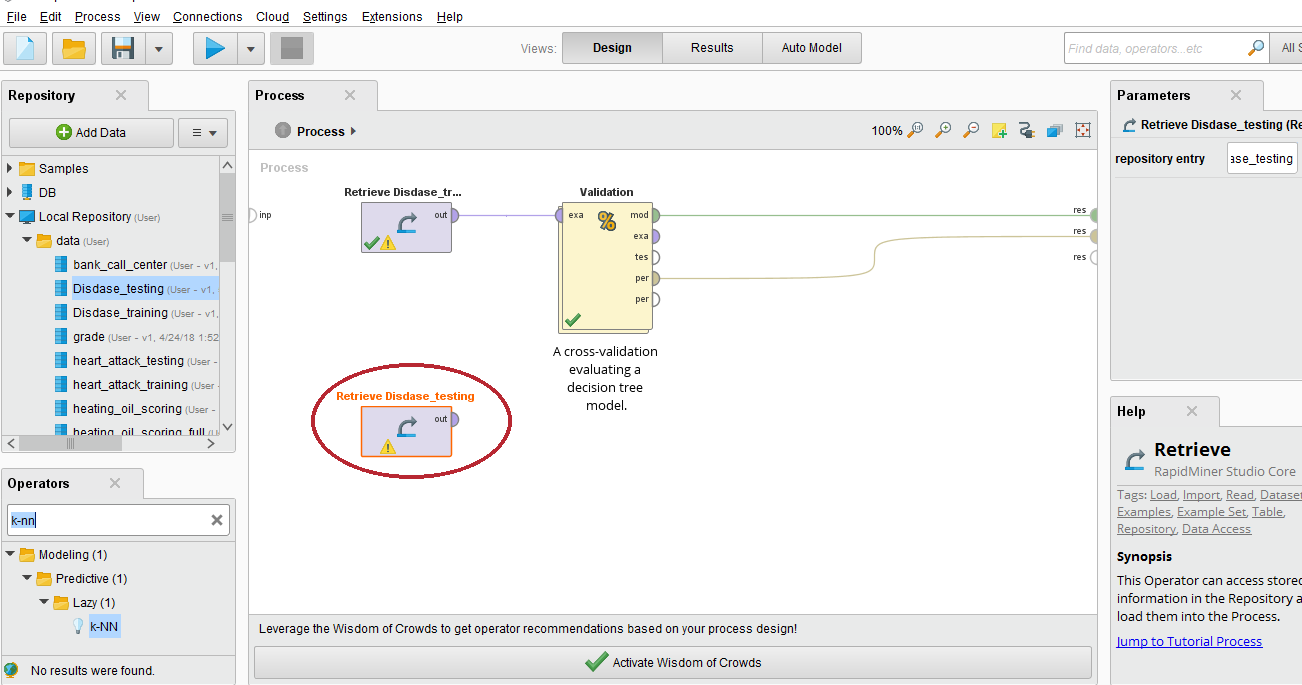
รูปที่ 4. คลิกขวา เลือก insert building block แล้วเลือก nominal cross validation

ir

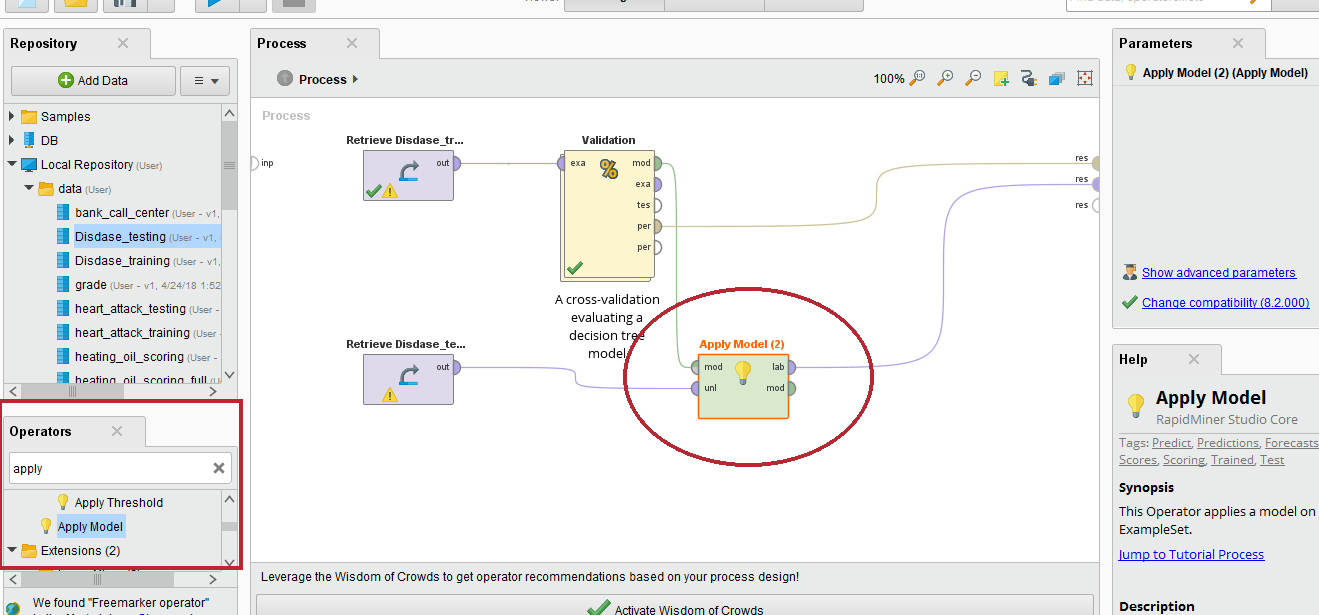
รูปที่ 5. ดับเบิ้ลคลิกเข้าใน validation เลือกใช้ K-NN ใส่ใน ช่อง training



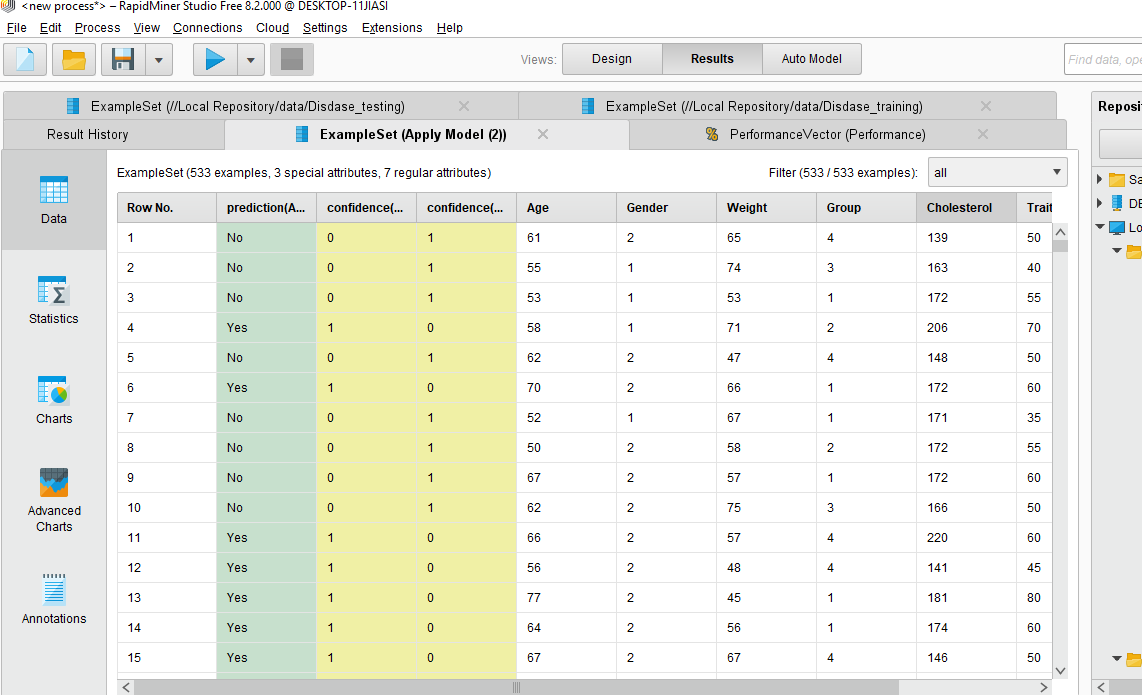
รูปที่ 6. ผลลัพธ์ที่ได้ออกมา



รูปที่ 7. เลือก Retrieve ใส่ไฟล์ disdase\_tasting.csv เพื่อมาคาดการณ์



รูปที่ 8. เลือก apply Model ใช้สำหรับ predict ข้อมูลใหม่



รูปที่ 9. ทำการรัน ผลลัพธ์ที่ออกมา จะพบว่ามีการคาดการณ์ว่าจะมีผู้สูงอายุที่เสี่ยงจะป่วยเป็นโรค กี่คน