# **Can you have heart disease?**

1. นายปิยังกูร แย้มทองดี,
2. นางสาวปุณยนุช แสงจันทร์,
3. นางสาวพิมพ์มาดา สมบูรณ์เศษ,
4. นายสุพัฒนนท์ แสงทอง

## **บทคัดย่อ**

ในอดีตโรคหลอดเลือดหัวใจมักเกิดในผู้สูงอายุ แต่ปัจจุบันเราพบว่าหนุ่มสาววัยทำงานนั้นมีความเสี่ยงเป็นโรคหลอดเลือดหัวใจสูงขึ้นกว่าแต่ก่อนมาก ซึ่งเป็นผลมาจากหลายปัจจัย โดยเฉพาะไลฟ์สไตล์การใช้ชีวิตที่เปลี่ยนไป เรื่องของโรคประจำตัวและพันธุกรรม ซึ่งหากใครมีสมาชิกในครอบครัวเป็นโรคหลอดเลือดหัวใจ หรือแม้แต่ตนเองเป็นโรคเบาหวาน หรือมีภาวะไขมันในเลือดสูง ความดันโลหิตสูง ก็เป็นการเพิ่มความเสี่ยงต่อการเป็นโรคหลอดเลือดหัวใจ ทางผู้จัดทำจึงมีการทำโมเดลหาความน่าจะเป็นที่จะเกิดโรคหัวใจเพื่อที่จะได้ทำการตรวจสอบหรือรักษาแต่เนิ่น ๆ

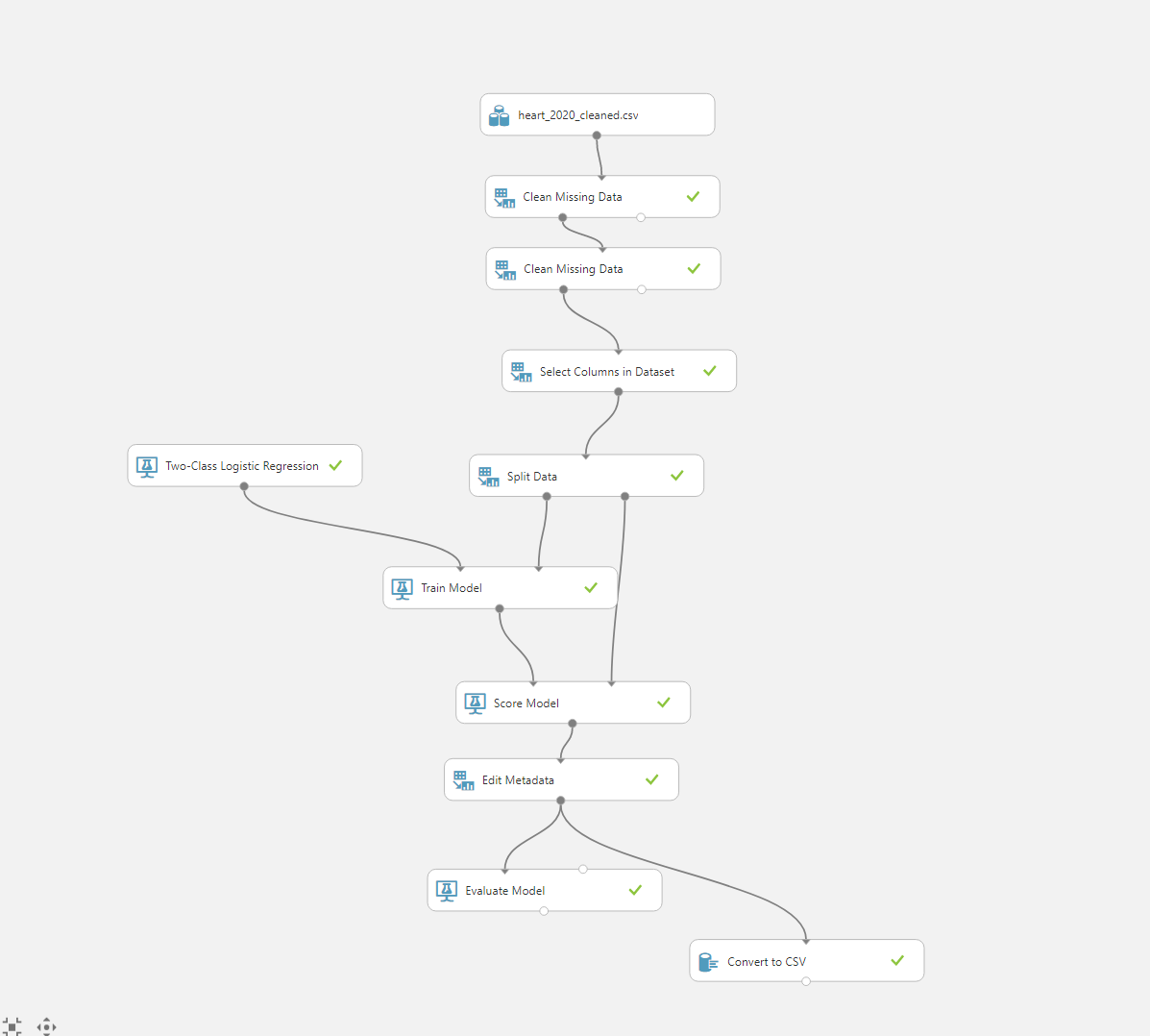
**คำหลัก : BMI(ค่าดัชนีมวลกาย), Heart Disease, Smoking, Stroke, Diabetic**

## **1.ที่มาและปัญหา**

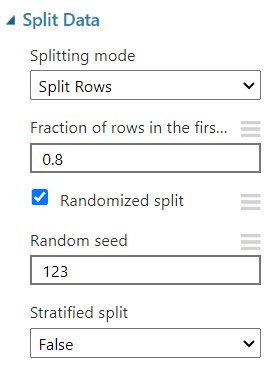
จากข้อมูลของCDCโรคหัวใจเป็นสาเหตุการตายอันดับต้น ๆ ของคนส่วนใหญ่ในสหรัฐอเมริกา (ชาวแอฟริกันอเมริกัน ชาวอเมริกันอินเดียน ชาวอะแลสกา และคนผิวขาว) ประมาณครึ่งหนึ่งของชาวอเมริกันทั้งหมด (47%) มีปัจจัยเสี่ยงอย่างน้อย 1 ใน 3 ประการของโรคหัวใจ ได้แก่ ความดันโลหิตสูง คอเลสเตอรอลสูง และการสูบบุหรี่ ตัวบ่งชี้ที่สำคัญอื่น ๆ ได้แก่ สถานะโรคเบาหวาน โรคอ้วน (ค่าดัชนีมวลกายสูง) ออกกำลังกายไม่เพียงพอ หรือดื่มแอลกอฮอล์มากเกินไป การตรวจหาและป้องกันปัจจัยที่มีผลกระทบต่อโรคหัวใจเป็นสิ่งสำคัญมากในการดูแลสุขภาพ ในทางกลับกัน การพัฒนาด้านคอมพิวเตอร์ช่วยให้สามารถประยุกต์ใช้วิธีการเรียนรู้ของเครื่องเพื่อตรวจหา "รูปแบบ" จากข้อมูลที่สามารถทำนายอาการของผู้ป่วยได้

**2.วิธีการ**

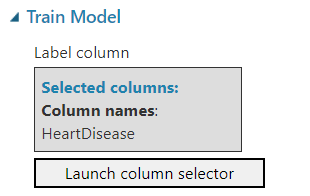
ใช้โมเดล Two-Class Logistic Regression เป็นการสร้างแบบจำลองแบบทำนายอนาคต ที่ได้จากข้อมูลเก่า



เลือกข้อมูลที่จำเป็น และแบ่งมาtrain 80% และ test 20%



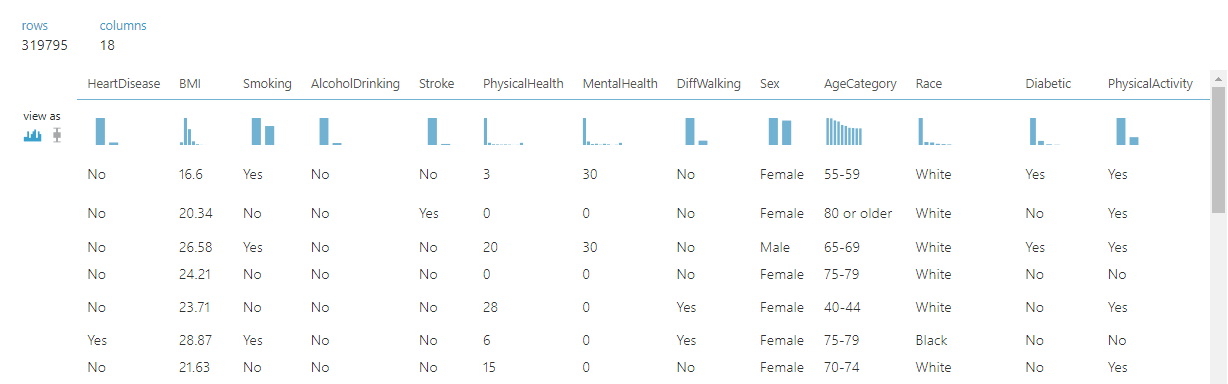
เราจะทำการtrain เพื่อหา heartdisease



**3.การออกแบบการทดลอง**

**3.1ชุดข้อมูลที่ใช้**

https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease



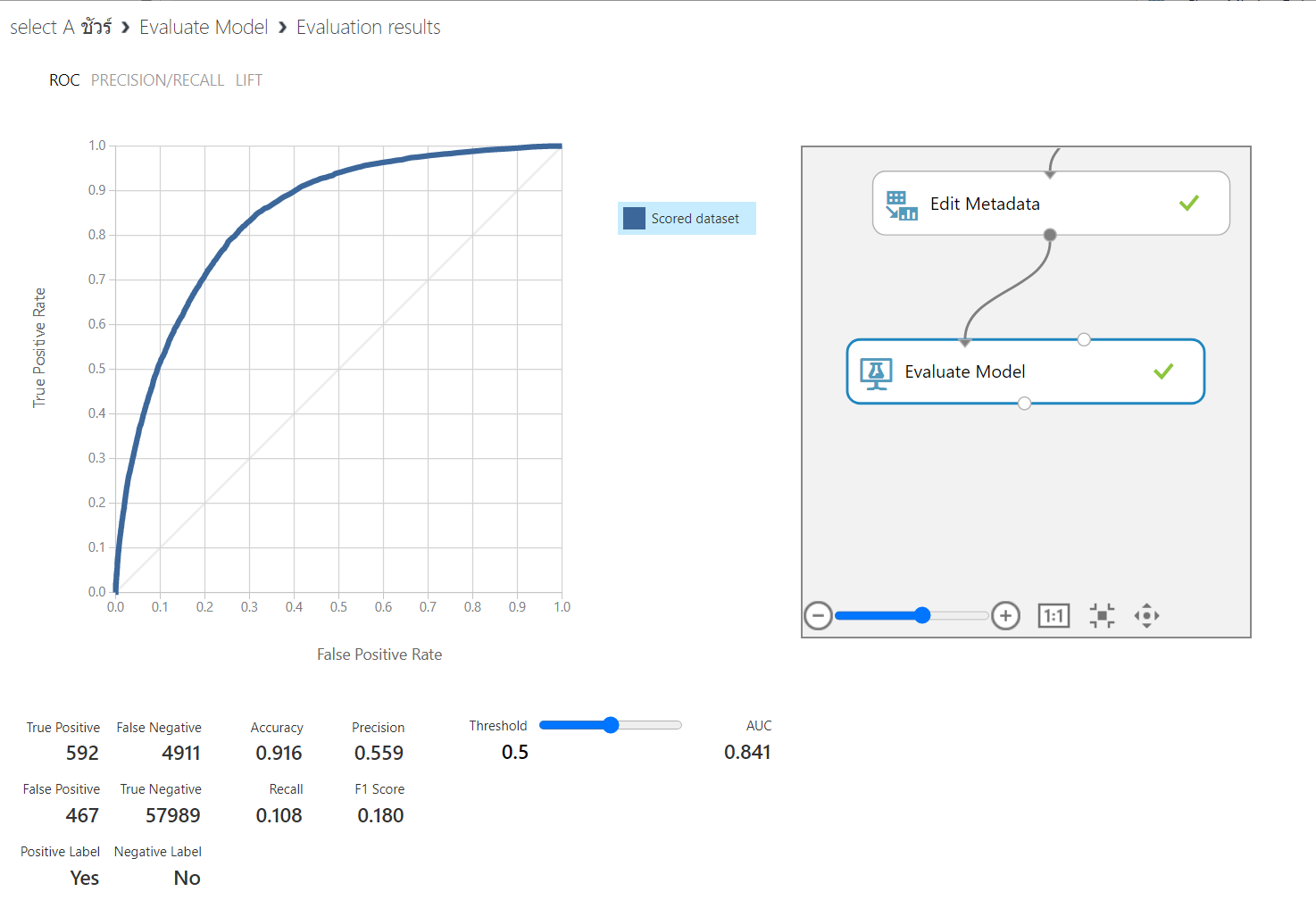
**3.2วิธีวัดความถูกต้องใช้**

ข้อมูลที่ได้ทำการวิเคราะห์มีดังนี้

* HeartDisease: โรคหัวใจ
* BMI:ค่าดัชนีมวลกาย
* Smoking:สูบบุหรี่
* AlcoholDrinking:การดื่มเครื่องดื่ม

แอลกอฮอล์

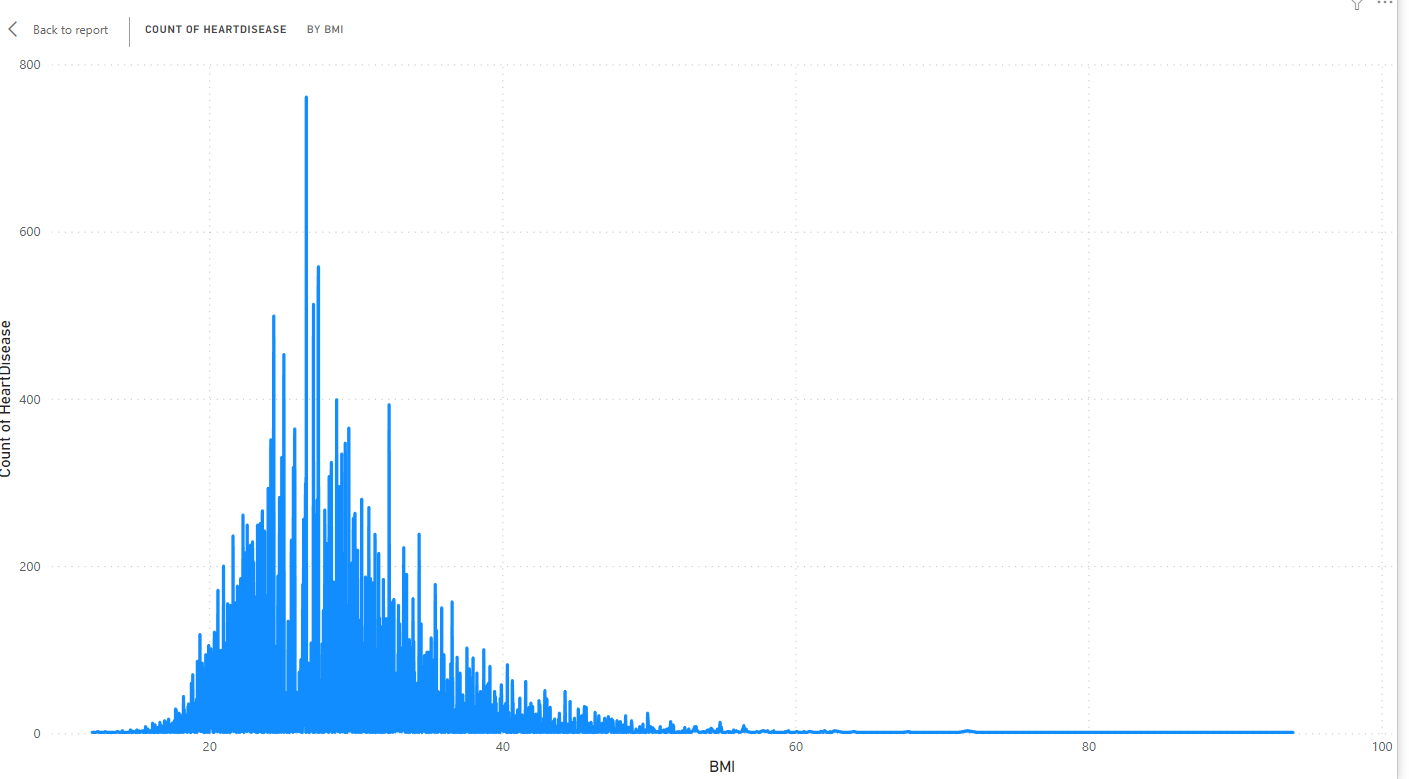
* Stroke:โรคหลอดเลือดสมอง
* PhysicalHealth:สุขภาพกาย
* MentalHealth:สุขภาพจิต
* DiffWalking: เดินแบบ diff
* Sex:เพศ
* AgeCategory:หมวดหมู่อายุ
* Race:สีผิว
* Diabetic:เบาหวาน
* PhysicalActivity:การออกกำลังกาย
* GenHealth: สุขภาพโดยทั่วไป
* SleepTime:เวลานอน
* Asthma:โรคหอบหืด
* KidneyDisease:โรคไต
* SkinCancer:มะเร็งผิวหนัง

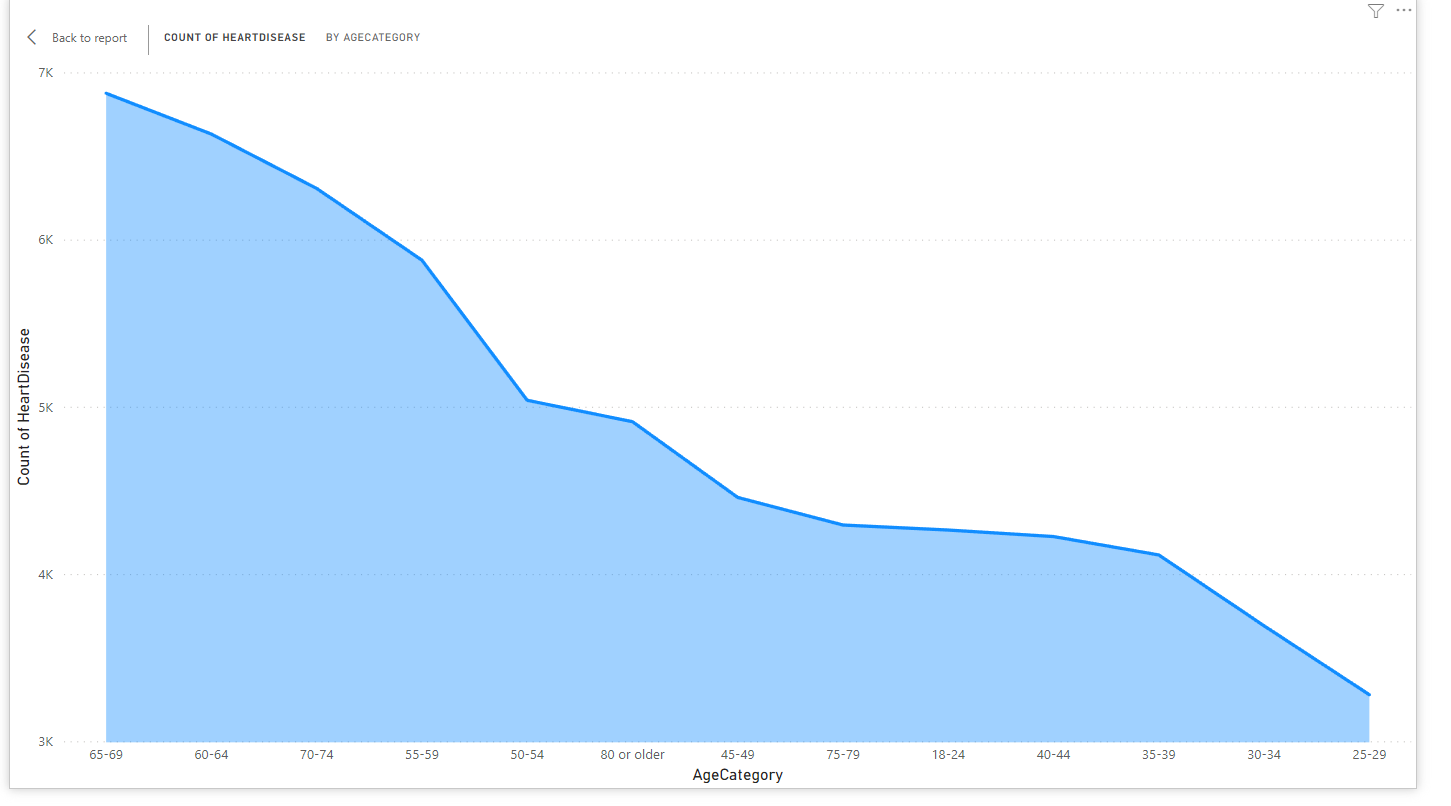


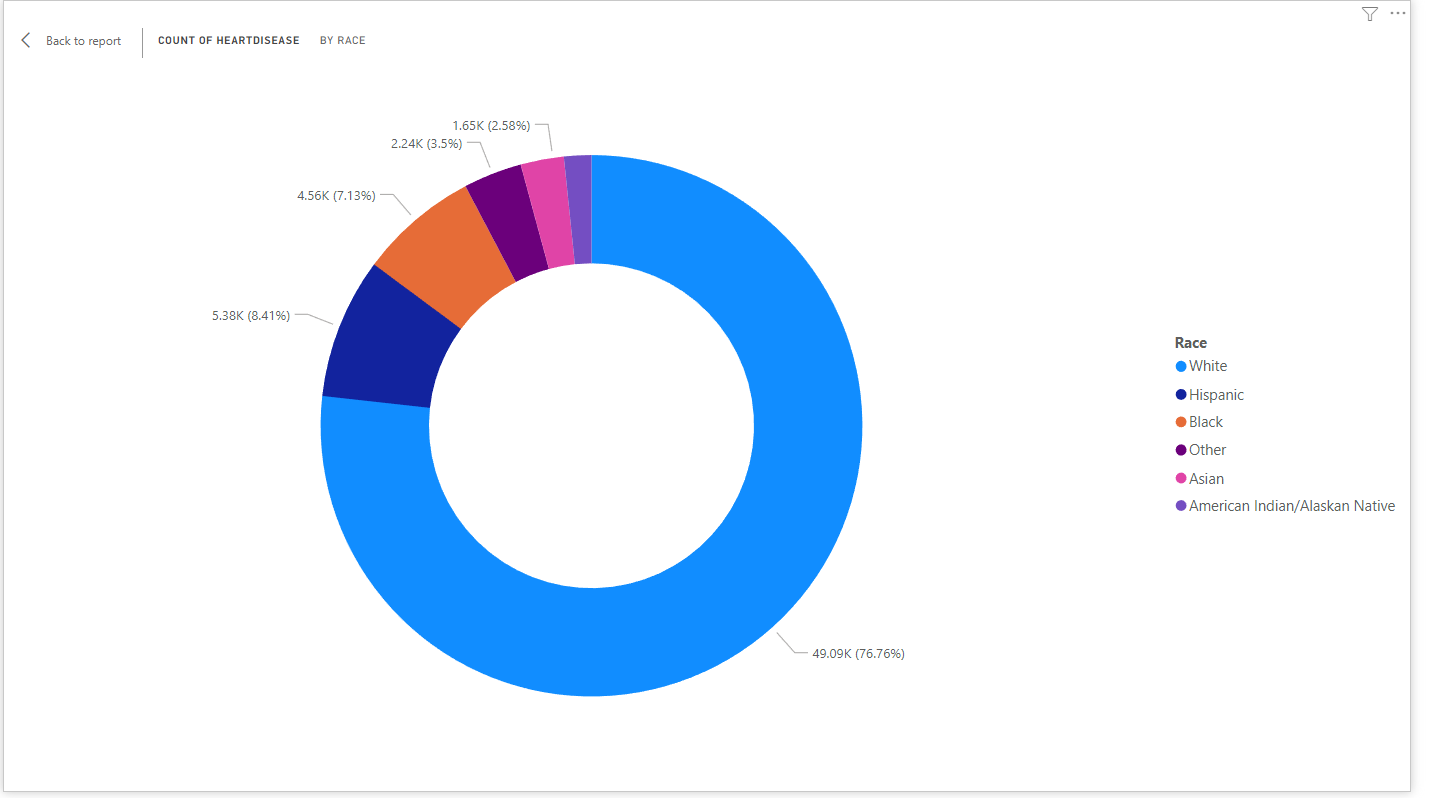
เมื่อเรา Visualization เส็ดเรียบร้อย เราจะได้ กราฟ roc มา ยิ่งกราฟออกห่างจากเส้น 0 มากเท่าไหร่แสดงว่า model นั้นยิ่งดี และค่า AUC ยิ่งเข้าใกล้ 1 เม่าไหร่ ยิ่ง

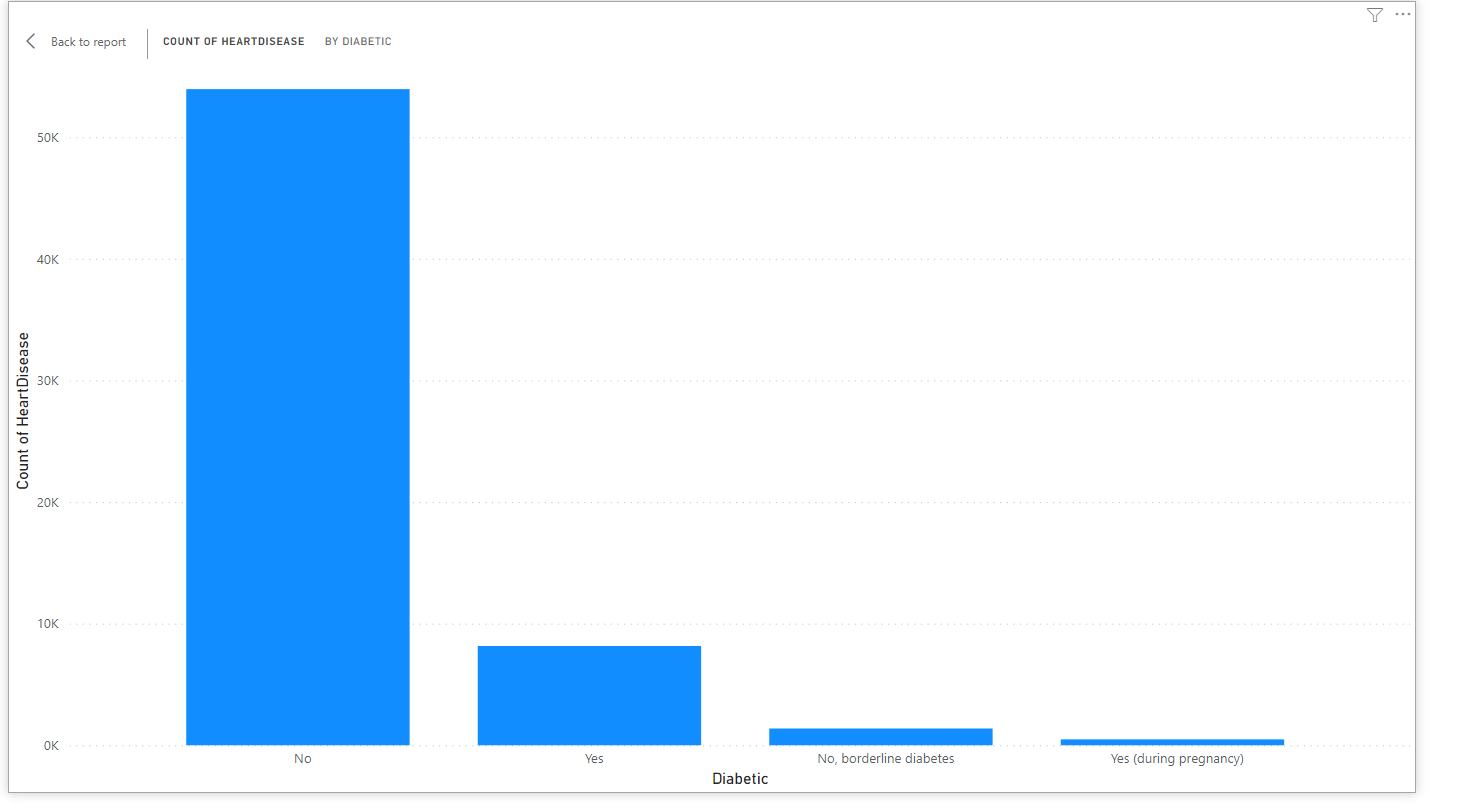
เราจะเห็นว่า ค่า AUC ของเรานั้น มีค่า อยู่ 0.841 หรือ 84%

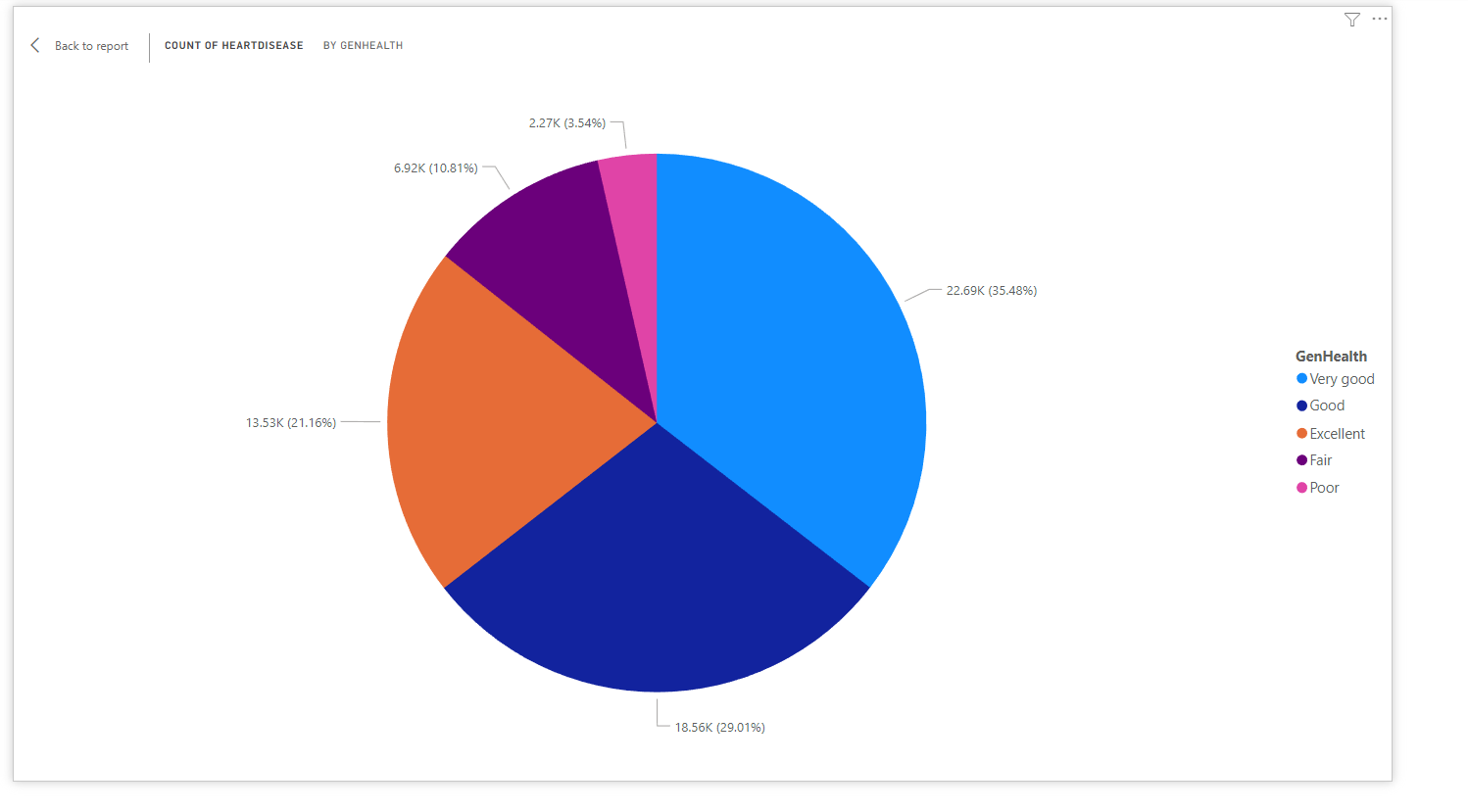
**4.ผลการทดลอง**

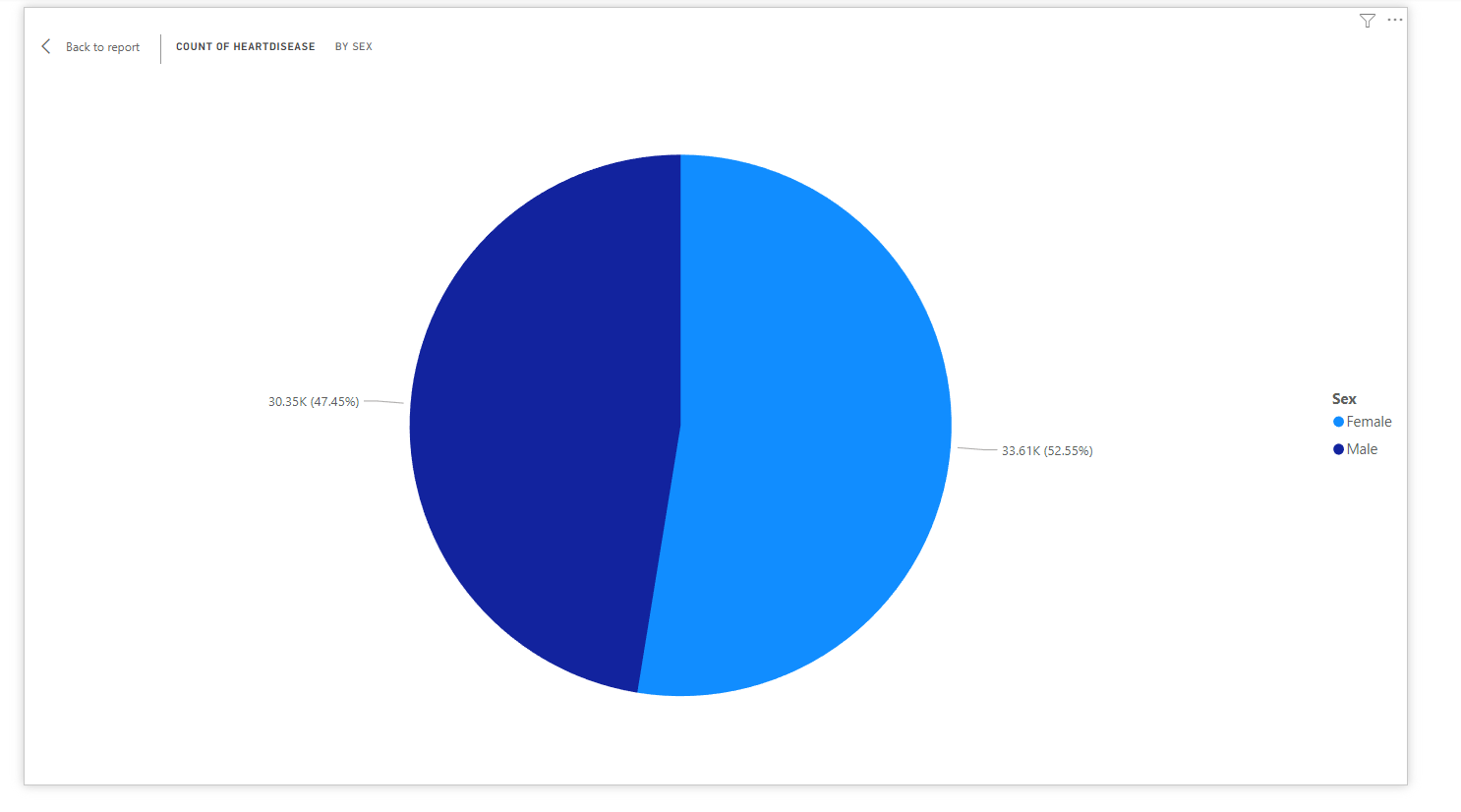


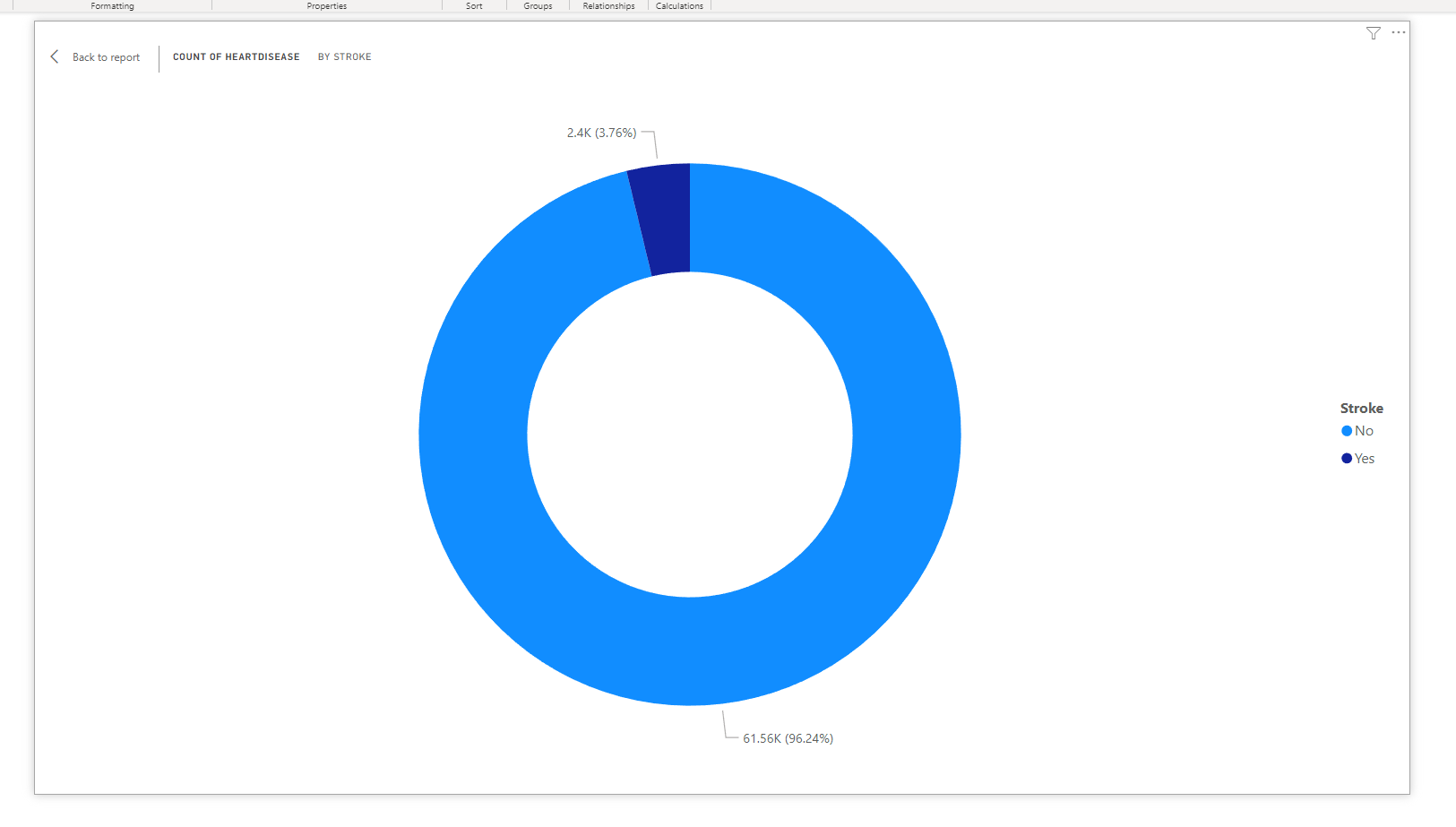












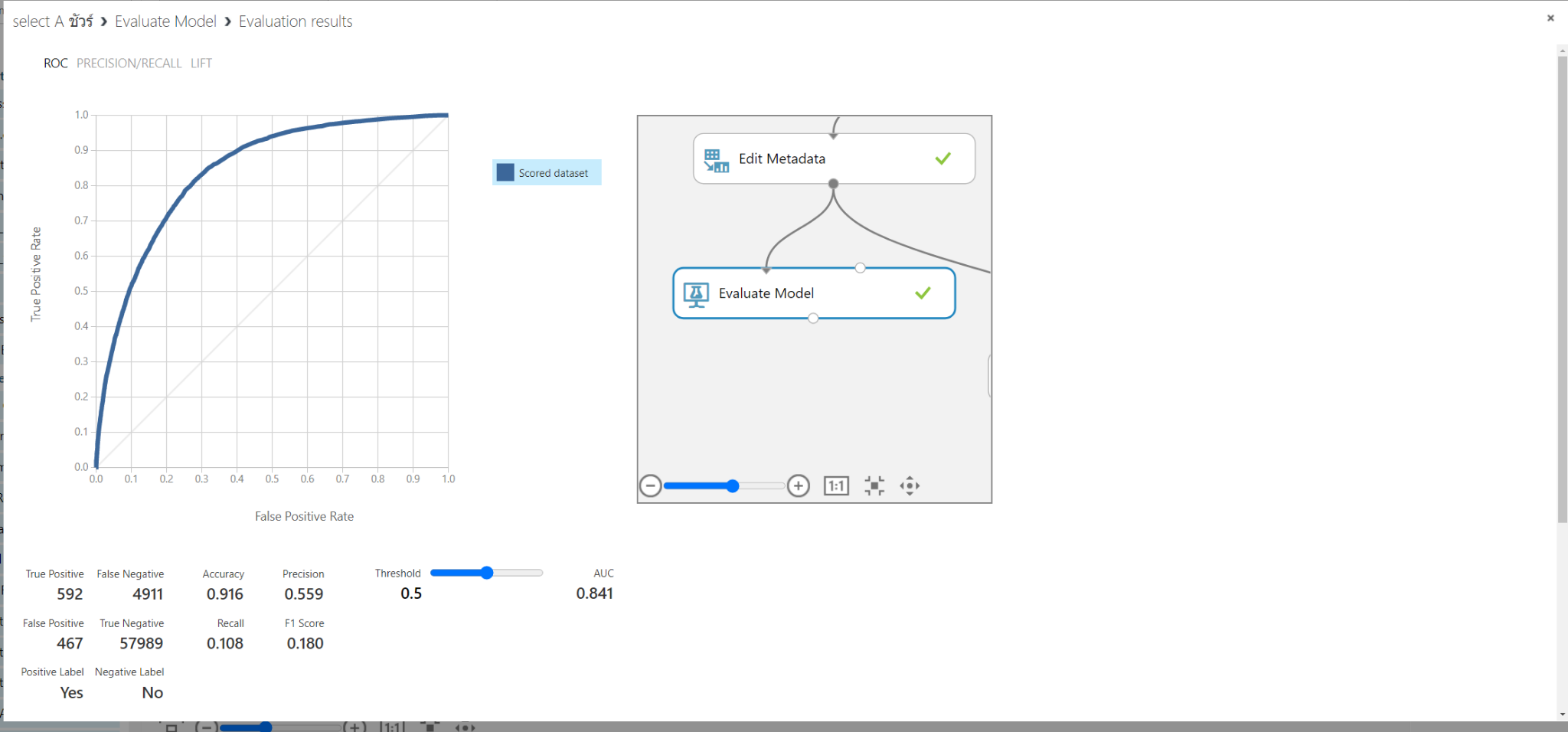
**5.สรุปผลการทดลอง**

จากข้อมูลการ Train 80%(255836 ข้อมูล)

และ Test 20%(63959 ข้อมูล)

มีความแม่นยำ Accuracy(91%) และ

ค่า true positive และ true negative มีค่า มากกว่า ค่า false positive และ false negative มีค่า AUC ถึง 0.841 และ มี กราฟ roc ห่างจากเส้นตรงกลางข้อนข้างมาก



จากการทดลองโมเดลของเรานั้นสามารถแก้ปัญหาผู้ป่วยที่มีความเสี่ยงของโรคภาวะหัวใจล่มเหลวได้ เพียงแค่กรอกข้อมูลเพียงเล็กน้อย ก็สามารถประเมินความเสี่ยงของโรคได้โดยไม่ต้องไปเสียเงินตรวจสุขภาพราคาแพง

และค่าแม่นยำ กว่า 90%ซึ่งทำให้Model ที่

ได้นั้น มีประโยชน์ต่อ ผู้ป่วยทำให้ผู้ป่วยสามารถรู้ตัวว่าเสี่ยงต่อโรคและรักษาได้ทันท่วงที

**6.เอกสารอ้างอิง**

* <https://www.kaggle.com/datasets/michaelbryantds/tornadoes>
* https://www.paolohospital.com/th-TH/center/Article