



รายงาน

การวิเคราะห์และเปรียบเทียบแบบของข้อมูลโรคเบาหวาน

จัดทำโดย

นาย ภคนรร์ห์ ตันติวุฒิ 6309680061

เสนอ

ผู้ช่วยศาสตราจารย์ ดร.ปกรณ์ลีสุทธิพรชัย

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา คพ.377

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ภาคการศึกษาที่ 2 ปีการศึกษา 2567

บทนำ

โรคเบาหวานเป็นโรคเรื้อรังที่มีอิทธิพลต่อสุขภาพของมนุษย์ในทุกช่วงวัย โรคนี้เกิดจากการเกิดผลข้างเคียงของฮอร์โมนอินซูลินภายในร่างกาย ทำให้ร่างกายไม่สามารถใช้อินซูลินได้อย่างเพียงพอ ซึ่งจะทำให้ระดับน้ำตาลในเลือดเพิ่มขึ้น ซึ่งเมื่อไม่ได้รักษาอย่างเหมาะสม จะสามารถเกิดภาวะแทรกซ้อนที่ร้ายแรงได้ เช่น โรคหัวใจและหลอดเลือด อัมพาต ไตวาย และการสูญเสียสายตา

การรักษาโรคเบาหวานทำได้โดยการควบคุมระดับน้ำตาลในเลือด การออกกำลังกาย และการรักษาทางยา เพื่อลดความเสี่ยงในการเกิดภาวะแทรกซ้อน อีกทั้งยังสามารถป้องกันโรคได้ด้วยการรักษาโปร่งใส และการดูแลสุขภาพที่ดี

การเข้าใจเกี่ยวกับโรคเบาหวาน เหตุผลที่เกิด อาการ และวิธีการรักษาที่เหมาะสม เป็นสิ่งสำคัญที่จะช่วยลดความเสี่ยงในการเป็นโรคและลดการเกิดภาวะแทรกซ้อนที่อาจเกิดขึ้นได้อย่างมีประสิทธิภาพ

คำอธิบายเกี่ยวกับข้อมูล

ข้อมูลสุขภาพทั้งหมด 392 ข้อมูล 8 ตัวแปร ดังนี้

Pregnancies : การตั้งครรภ์(ครั้ง)

การตั้งครรภ์เป็นครั้งแรก (Primiparity) การตั้งครรภ์ครั้งแรกอาจเป็นปัจจัยเสี่ยงที่เกี่ยวข้องกับเบาหวานในอนาคต การตั้งครรภ์ครั้งแรกในวัยที่มากกว่า 25 ปีอาจมีความเสี่ยงที่สูงขึ้นสำหรับโรคเบาหวานชนิดที่ 2 ในอนาคต

จำนวนครั้งการตั้งครรภ์ (Number of Pregnancies) การตั้งครรภ์หลายๆ อาจมีผลต่อการเปลี่ยนแปลงฮอร์โมนและการควบคุมน้ำตาลในเลือดซึ่งเป็นสาเหตุที่ทำให้เกิดเบาหวานในอนาคต

Glucose : กลูโคส (mg/dL)

ระดับกลูโคส (mg/dL)	ความหมาย
น้อยกว่า 70	ต่ำกว่าปกติ
70 - 99	ปกติ
100 - 125	ภาวะเสี่ยงเป็นเบาหวาน
126 ขึ้นไป	เสี่ยงเป็นโรคเบาหวาน

หมายเหตุ: ค่าที่แสดงเป็นข้อมูลเพียงทั่วไป และอาจมีความแตกต่างไปตามคู่มือการวินิจฉัยที่ใช้ งาน กรุณาตรวจสอบกับแพทย์หรือผู้เชี่ยวชาญเพื่อคำแนะนำที่ถูกต้องตามสถานการณ์ของแต่ละบุคคล

BloodPressure : ความดันโลหิต (mmHg)

ระดับความดันเลือด (mmHg)	ความหมาย
น้อยกว่า 120/80	ปกติ
120/80 - 139/89	ภาวะความดันโลหิตสูงเล็กน้อย (ปกติเล็กน้อย)
140/90 ขึ้นไป	เสี่ยงเป็นโรคความดันโลหิตสูง (ภาวะความดันโลหิตสูง)

SkinThickness : ความหนาของผิวหนัง (mm)

ปกติ : ความหนาของผิวหนังอยู่ในเกณฑ์ปกติสำหรับบุคคลในกลุ่มที่ตรวจวัด

ภาวะความเสี่ยง : ความหนาของผิวหนังมีค่าสูงเล็กน้อย ซึ่งอาจแสดงถึงความเสี่ยงต่อการเป็นโรคเบาหวาน แต่ไม่ใช่สิ่งที่สามารถวินิจฉัยโรคได้โดยตรง

เสี่ยงเป็นโรคเบาหวาน : ความหนาของผิวหนังมีค่าสูง ซึ่งอาจเป็นสัญญาณให้ทราบว่ามีความเสี่ยงต่อการเป็นโรคเบาหวานได้ และอาจต้องการการตรวจวินิจฉัยเพิ่มเติมเพื่อการรักษาหรือการสแกนโรค

Insulin : ฮอร์โมนที่ผลิตโดยเซลล์เบต้าในหลอดเลือดทางเดินอาหารในส่วนที่เรียกว่าไขมันในท้อง (mu U/ml)

ระดับ	คำอธิบาย
ปกติ	ระดับ Insulin อยู่ในเกณฑ์ปกติ
ภาวะความเสี่ยง	ระดับ Insulin สูงขึ้นแสดงถึงความเสี่ยงต่อการเป็นโรคเบาหวาน
เสี่ยงเป็นโรคเบาหวาน	ระดับ Insulin สูงมากแสดงถึงความเสี่ยงที่สูงต่อการเป็นโรคเบาหวาน

สำหรับระดับ Insulin ไม่สามารถระบุค่าที่แน่นอนได้โดยไม่มีการตรวจวัด แต่สามารถบอกได้ว่ามีระดับ Insulin ที่สูงขึ้นอาจแสดงถึงความเสี่ยงต่อการเป็นโรคเบาหวานมากขึ้น ดังนั้นหากมีค่า Insulin ที่สูงขึ้นควรรีบพบแพทย์หรือผู้เชี่ยวชาญด้านสุขภาพเพื่อประเมินความเสี่ยงและการดูแลสุขภาพให้เหมาะสม

BMI : ดัชนีมวลกาย

ช่วง BMI	สถานะ
น้อยกว่า 18.5	น้ำหนักน้อยกว่าเกณฑ์ (Underweight)
18.5 - 24.9	น้ำหนักปกติ (Normal weight)
25 - 29.9	น้ำหนักเกิน (Overweight)
30 ขึ้นไป	อ้วน (Obese)

BMI เกี่ยวข้องกับโรคเบาหวานในทางทฤษฎีซึ่งสามารถใช้เป็นตัวบ่งชี้ความเสี่ยงต่อโรคนี้ได้ คนที่มี BMI อยู่ในระดับที่สูงกว่าเกณฑ์มาตรฐานอาจมีความเสี่ยงต่อการเป็นโรคเบาหวานสูงขึ้น เนื่องจากการมีน้ำหนักที่เกินเป้าหมายส่งผลต่อการดูแลสุขภาพที่ไม่ดี ซึ่งอาจเป็นอันตรายต่อการเกิดโรคเบาหวานในระยะยาว

DiabetesPedigreeFunction(DPF) : ค่าที่ใช้ในการประเมินความเสี่ยงของโรคเบาหวาน

มีสาเหตุมาจากพ่อแม่ หรือภูมิพันธุกรรม ซึ่งมักจะนำมาใช้ในการคำนวณความเสี่ยงของการเป็นโรคเบาหวานในระบบต่าง ๆ และส่วนใหญ่จะมีค่าอยู่ระหว่าง 0 ถึง 2

ค่า DPF สูงขึ้นมักจะแสดงถึงความเสี่ยงที่สูงขึ้นในการเป็นโรคเบาหวาน โดยเฉพาะหากมีประวัติความเสี่ยงในครอบครัว การวัดค่า DPF สามารถช่วยในการคาดการณ์ความเสี่ยงของการเป็นโรคเบาหวานในอนาคตได้

Age : อายุ

อายุ ของบุคคลที่ถูกวัดหรือบันทึกลงในชุดข้อมูล อายุเป็นตัวแปรที่สำคัญในการประเมินความเสี่ยงของโรคเบาหวาน เนื่องจากโรคนี้มักจะพบบ่อยในกลุ่มผู้สูงอายุและมีความเสี่ยงเพิ่มขึ้นตามอายุ

Outcome : โรคเบาหวาน

0 คือ ผู้ที่ไม่ได้เป็นโรคเบาหวาน

1 คือ ผู้ที่เป็นโรคเบาหวาน

อาการของผู้ที่เป็นโรคเบาหวาน Diabetes

- ปัสสาวะบ่อย: การต้องปัสสาวะบ่อยเกินไป เป็นสัญญาณที่มักพบในผู้ป่วยเบาหวาน เนื่องจากน้ำตาลในเลือดสูง ทำให้ไปตกลงในปัสสาวะ
- กระหายน้ำมาก: ความกระหายน้ำมากเกิดจากการต้องดื่มน้ำมากเพื่อชดเชยการสูญเสียน้ำที่มีผลจากการปัสสาวะบ่อย
- ร่างกายเหนื่อยง่าย: การใช้พลังงานไม่เพียงพอ เนื่องจากเซลล์ไม่สามารถใช้น้ำตาลเป็นพลังงานได้อย่างเต็มที่
- เจ็บป่วยง่าย: เนื่องจากน้ำตาลสูงในเลือดอาจทำให้ร่างกายมีภูมิคุ้มกันต่ำลง

- ริมฝีปากและผิวหนังมีด: ผู้ป่วยบางคนอาจมีปัญหาริมฝีปากหรือบริเวณผิวหนังที่มืดขึ้น
- เจ็บท้อง: อาจมีอาการเจ็บบริเวณท้องเนื่องจากการอาจทำให้เกิดภาวะเสี่ยงต่อการติดเชื้อในกระเพาะอาหาร

โรคเบาหวานมักมีอาการไม่ชัดเจนในระยะแรก และบางครั้งอาจไม่มีอาการเลย การตรวจเบาหวานอย่างสม่ำเสมอเป็นสำคัญสำหรับผู้ที่มีปัจจัยเสี่ยง โดยการรักษาโดยเจตนาและการดูแลรักษาเป็นปัจจัยสำคัญในการควบคุมโรคให้อยู่ในสภาวะที่ดี

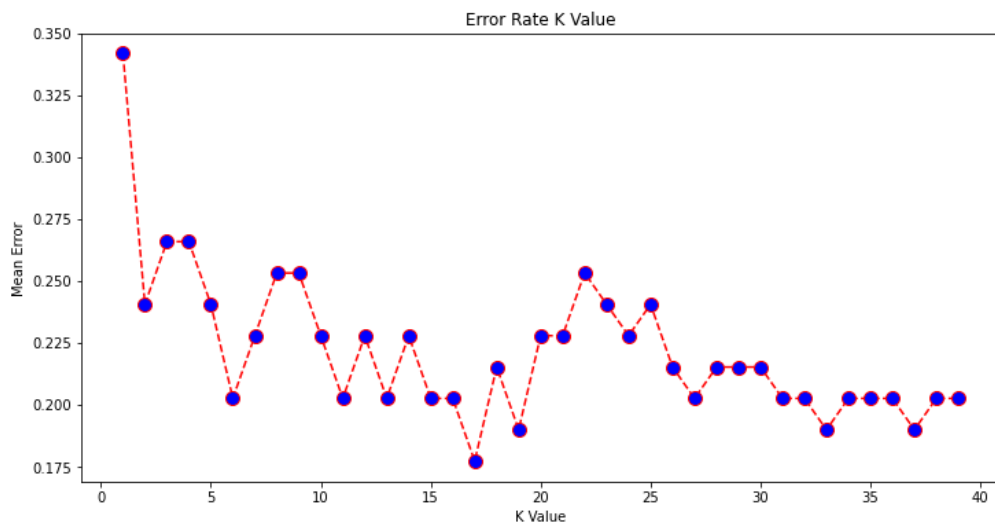
หลักการและขั้นตอน การสกัด การเลือก และการเตรียมลักษณะ

- จากข้อมูลที่น่ามานั้น มีข้อมูลที่เป็น Missing Value จากนั้นทำการลบแถวที่มีค่าเป็น 0
- แบ่งข้อมูล Train set : Test set เป็น 80:20 คือ 314 ข้อมูลและ 78 ข้อมูล ตามลำดับ

เทคนิคที่ใช้

KNN

1. ทำ Cross validation (cv=10) ได้ accuracy = 0.71 และ standard deviation = 0.06
2. คำนวณค่า error สำหรับค่า K ระหว่าง 1 - 40 ซึ่งค่า K ที่ให้ error น้อยที่สุด คือ K=14



3. สร้าง Confusion Matrix และ Performance Metrics

[[45 9] [10 15]]				
	precision	recall	f1-score	support
0	0.82	0.83	0.83	54
1	0.62	0.60	0.61	25
accuracy			0.76	79
macro avg	0.72	0.72	0.72	79
weighted avg	0.76	0.76	0.76	79

Naïve Bays

1. ทำ Cross validation (cv=10) ของ Gaussian ได้ค่า accuracy = 0.86 และ standard deviation = 0.08
และ Multinomial ได้ค่า accuracy = 0.85 และ standard deviation = 0.00
และ BernoulliNB ได้ค่า accuracy = 0.85 และ standard deviation = 0.01

2. สร้าง Confusion Matrix และ Performance Metrics พบว่า ตัวแบบ Gaussian ดีกว่า Multinomial และ Bernoulli สำหรับ การรันครั้งนี้และเมื่อลองนำ มาเปรียบเทียบกับ การแบ่ง Train และ Test ได้ผลที่ไม่ overfitting

```
Confusion Matrix for GaussianNB:
[[47  7]
 [ 7 18]]

Classification Report for GaussianNB:
              precision    recall  f1-score   support

     0       0.87         0.87         0.87         54
     1       0.72         0.72         0.72         25

 accuracy          0.82
 macro avg         0.80         0.80         0.80         79
weighted avg         0.82         0.82         0.82         79
```

Gaussian

```
Classification Report for MultinomialNB:
              precision    recall  f1-score   support

     0       0.68         1.00         0.81         54
     1       0.00         0.00         0.00         25

 accuracy          0.68
 macro avg         0.34         0.50         0.41         79
weighted avg         0.47         0.68         0.56         79
```

Multinomial

```
Confusion Matrix for BernoulliNB:
[[53  1]
 [25  0]]

Classification Report for BernoulliNB:
              precision    recall  f1-score   support

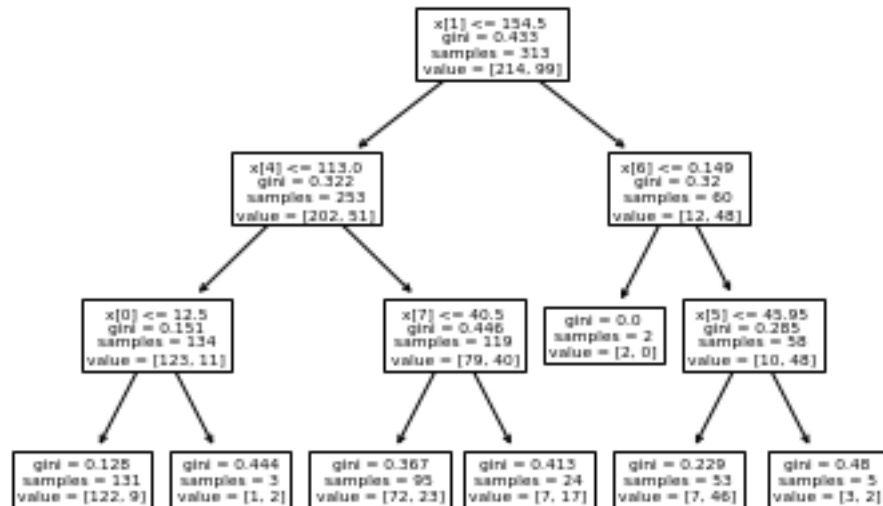
     0       0.68         0.98         0.80         54
     1       0.00         0.00         0.00         25

 accuracy          0.67
 macro avg         0.34         0.49         0.40         79
weighted avg         0.46         0.67         0.55         79
```

BernoulliNB

Tree

1. ทำ Cross validation (cv=10) ได้ accuracy = 0.72 และ standard deviation = 0.07
2. ให้ค่า max depth = 3 และใช้ GINI ในการคำนวณ



3. สร้าง Confusion Matrix และ Performance Metrics

```
[[44  4]
 [18 13]]

Classification Report:

```

	precision	recall	f1-score	support
0	0.71	0.92	0.80	48
1	0.76	0.42	0.54	31
accuracy			0.72	79
macro avg	0.74	0.67	0.67	79
weighted avg	0.73	0.72	0.70	79

การประเมินประสิทธิภาพของแบบจำลอง

Model	Train score	Test score	Diff
K-Neighbors Classifier	82	81	1
Gaussian Naïve Bayes	76	82	6
Tree	84	77	7

จากการเปรียบเทียบทั้ง 3 แบบ คือ K-Neighbors Classifier, Gaussian Naïve Bayes และTree ในการพยากรณ์ข้อมูลโรคเบาหวาน พบว่า Tree มีโอกาส เกิดปัญหา Overfitting อยู่บ้าง และจะเห็นได้ว่า โมเดล K-Neighbors Classifier ให้ประสิทธิภาพดี จึงเป็นทางเลือกที่ดีที่สุดเนื่องจากมีความแม่นยำสูงและไม่มีการเกิด overfitting ต่างจากโมเดลที่ผ่านมา จึงทำให้เราเลือกโมเดล K-Neighbors Classifier

วิเคราะห์ผลลัพธ์ที่ได้

จากการวิเคราะห์ข้อมูลโรคเบาหวาน มีปัจจัยที่ส่งผลต่อโรคทั้งหมดอยู่ 7 ปัจจัย จำนวนชุดข้อมูลทั้งหมด 392 ข้อมูล และโมเดลที่ใช้จำลองคือ K-Neighbors Classifier ในส่วน Confusion Metrix ที่ได้จาก K-Neighbors Classifier พบว่า จากข้อมูลในการ Test ทั้งหมด 79 ข้อมูล ผู้ที่เป็นโรคเบาหวาน(Diabetes Mellitus) ตัวโมเดลทายถูกว่าเป็นโรคเบาหวาน 11 คน และทายถูกว่าไม่เป็นโรคเบาหวาน 47 คน คิดเป็น 73% และทายผิด 21 คน คิดเป็น 27%

โดยสรุปแล้วการได้รับการวินิจฉัยที่แม่นยำและรวดเร็ว เราสามารถรับมือกับโรคได้อย่างมั่นใจ นอกจากนี้เรายังสามารถลดความเสี่ยงที่อาจเกิดขึ้นในอนาคตได้ด้วย การรับรู้และการตอบสนองที่รวดเร็วสามารถช่วยให้การรักษาเป็นไปอย่างมีประสิทธิภาพและป้องกันภาวะที่รุนแรงขึ้นในอนาคตได้

Feature Selection

เมื่อทำ Feature Selection ทำให้ได้ข้อมูลดังนี้

```
(392, 8)
[0.10470769 0.23200443 0.08551994 0.10039505 0.12179629 0.10888611
 0.10614876 0.14054171]
(392, 2)
Selected features: Index(['Glucose', 'Age'], dtype='object')
```

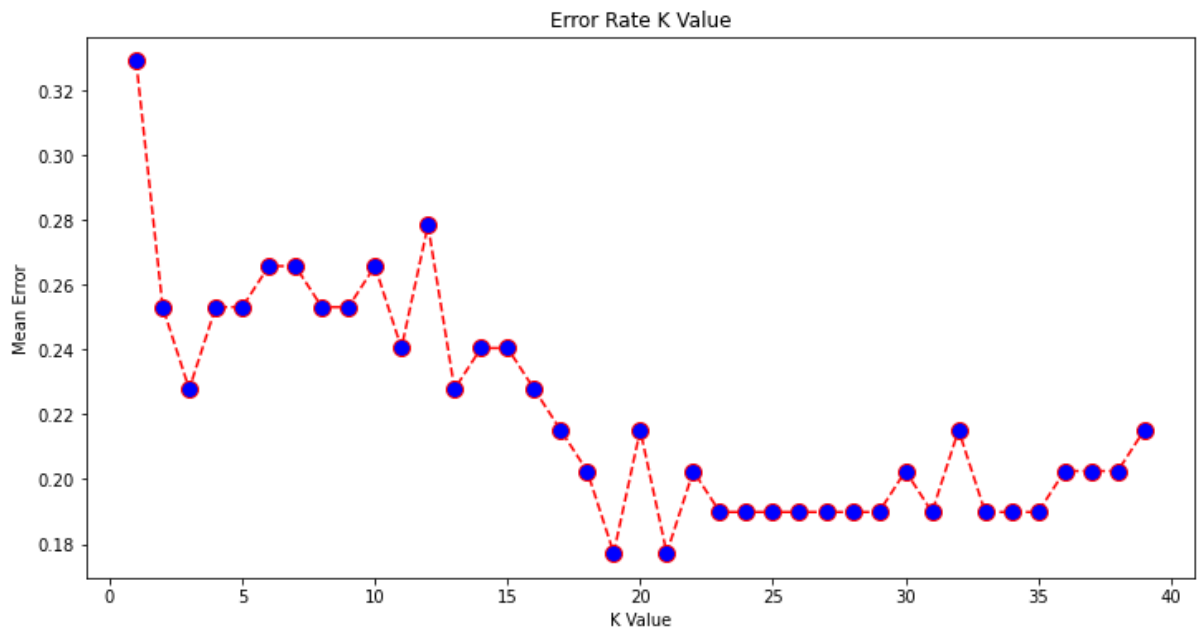
โมเดลได้ทำการเลือก Feature ที่มีความเกี่ยวข้องในคลาสที่น่าสนใจมา 2 Feature จากทั้งหมด 8 Feature ได้แก่ Glucose , Age

สรุปได้ว่า การวิเคราะห์ข้อมูลโดยใช้เทคนิค Feature Selection บนข้อมูลโรคเบาหวานพบว่า Glucose (น้ำตาลในเลือด) และ Age (อายุ) เป็นตัวแปรที่มีความสัมพันธ์กับโรคเบาหวานอย่างมีนัยสำคัญ โดยทั้งสองตัวแปรมีผลต่อการควบคุมระดับน้ำตาลในเลือดของร่างกาย โรคเบาหวานเป็นโรคที่มีความเสี่ยงเพิ่มขึ้นเมื่อมีระดับน้ำตาลในเลือดสูง และเมื่อเป็นไปได้ว่าระบบการควบคุมน้ำตาลในเลือดของร่างกายจะทำงานไม่เท่าทันเมื่อเทียบกับเวลาที่ยังเยาว์ ทำให้มีความน่าจะเป็นที่ระดับน้ำตาลในเลือดจะสูงขึ้นในกลุ่มคนที่มีอายุมากกว่า

Feature Selection และ เทคนิคที่ใช้

KNN

1. ทำ Cross validation (cv=10) ได้ accuracy = 0.75 และ standard deviation = 0.06
2. คำนวณค่า error สำหรับค่า K ระหว่าง 1 - 40 ซึ่งค่า K ที่ให้ error น้อยที่สุด คือ K=19 และ 21



3. สร้าง Confusion Matrix และ Performance Metrics

```
[[51  3]
 [17  8]]
```

	precision	recall	f1-score	support
0	0.75	0.94	0.84	54
1	0.73	0.32	0.44	25
accuracy			0.75	79
macro avg	0.74	0.63	0.64	79
weighted avg	0.74	0.75	0.71	79

Naïve Bays

1. ทำ Cross validation (cv=10) ของ Gaussian ได้ค่า accuracy = 0.77 และ standard deviation = 0.08
และ Multinomial ได้ค่า accuracy = 0.68 และ standard deviation = 0.00
และ BernoulliNB ได้ค่า accuracy = 0.67 และ standard deviation = 0.01
2. สร้าง Confusion Matrix และ Performance Metrics พบว่า ตัวแบบ Gaussian ดีกว่า Multinomial และ Bernoulli สำหรับ การรันครั้งนี้และเมื่อลองนำ มาเปรียบเทียบกับ การแบ่ง Train และ Test ได้ผลที่ไม่ overfitting
3. สร้าง Confusion Matrix และ Performance Metrics

```
Confusion Matrix for GaussianNB:
[[49  5]
 [13 12]]

Classification Report for GaussianNB:
              precision    recall  f1-score   support

     0       0.79         0.91         0.84         54
     1       0.71         0.48         0.57         25

 accuracy          0.77         0.77         0.77         79
 macro avg         0.75         0.69         0.71         79
 weighted avg      0.76         0.77         0.76         79
```

Gaussian

```
Confusion Matrix for MultinomialNB:
[[54  0]
 [25  0]]

Classification Report for MultinomialNB:
              precision    recall  f1-score   support

     0       0.68         1.00         0.81         54
     1       0.00         0.00         0.00         25

 accuracy          0.68         0.68         0.68         79
 macro avg         0.34         0.50         0.41         79
 weighted avg      0.47         0.68         0.56         79
```

Multinomial

```
Confusion Matrix for BernoulliNB:
[[53  1]
 [25  0]]

Classification Report for BernoulliNB:
              precision    recall  f1-score   support

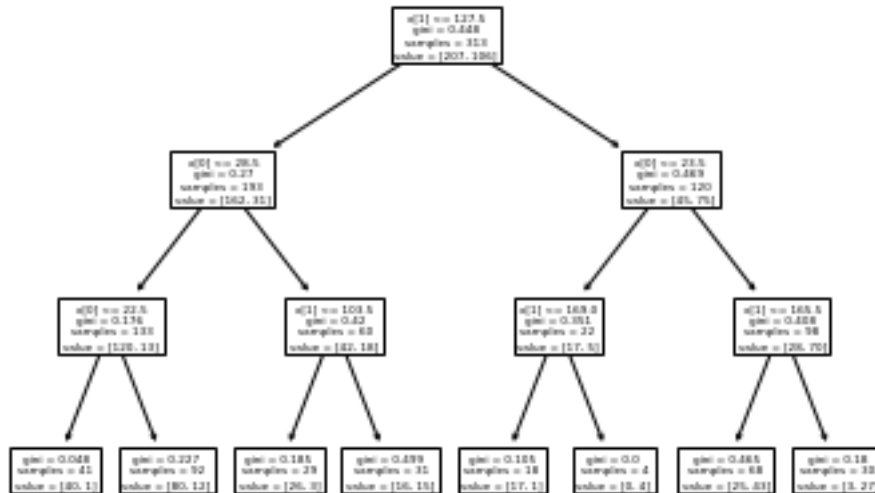
     0       0.68         0.98         0.80         54
     1       0.00         0.00         0.00         25

 accuracy          0.67         0.67         0.67         79
 macro avg         0.34         0.49         0.40         79
 weighted avg      0.46         0.67         0.55         79
```

BernoulliNB

Tree

1. ทำ Cross validation (cv=10) ได้ accuracy = 0.80 และ standard deviation = 0.07
2. ให้ค่า max depth = 3 และใช้ GINI ในการคำนวณ



3. สร้าง Confusion Matrix และ Performance Metrics

Confusion Matrix:					
[[44 11]					
[5 19]]					
Classification Report:					
	precision	recall	f1-score	support	
0	0.90	0.80	0.85	55	
1	0.63	0.79	0.70	24	
accuracy			0.80	79	
macro avg	0.77	0.80	0.77	79	
weighted avg	0.82	0.80	0.80	79	

การประเมินประสิทธิภาพของแบบจำลอง

	ทำ Feature Selection		
Model	Train score	Test score	Diff
K-Neighbors Classifier	78	87	9
Gaussian Naïve Bayes	79	77	2
Tree	80	79	1

จากการเปรียบเทียบทั้ง 3 แบบ คือ K-Neighbors Classifier, Gaussian Naïve Bayes และ Tree ในการ

พยากรณ์ข้อมูลโรคเบาหวาน พบว่า K-Neighbors Classifier มีโอกาส เกิดปัญหา Overfitting อยู่บ้าง เห็นว่าโมเดลอาจมีความ
ล่าช้าเมื่อใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน (Test set) อย่างไรก็ตาม คะแนน Test score ยังคงสูงอยู่ที่ 87% ซึ่งหมายความว่า
โมเดลยังมีประสิทธิภาพที่ดีในการทำนาย และจะเห็นได้ว่า โมเดล Tree ให้ประสิทธิภาพดี จึงเป็นทางเลือกที่ดีที่สุดเนื่องจากมี
ความแม่นยำสูงและไม่มีการเกิด overfitting ต่างจากโมเดลที่ผ่านมา จึงทำให้เราเลือกโมเดล Tree ในการพยากรณ์ข้อมูล

ตารางเปรียบเทียบระหว่างการทำ Feature Selection และไม่ทำ Feature Selection

	ทำ Feature Selection			ไม่ทำ Feature Selection		
Model	Train score	Test score	Diff	Train score	Test score	Diff
K-Neighbors Classifier	78	87	9	82	81	1
Gaussian Naïve Bayes	79	77	2	76	82	6
Tree	80	79	1	84	77	7

จากตารางการเปรียบเทียบข้อมูลระหว่างการทำ Feature Selection หรือ ไม่ทำ Feature Selection มีผลต่อ
ประสิทธิภาพของโมเดลอย่างชัดเจน หาก ทำ Feature Selection จะเหมาะกับ Gaussian Naïve Bayes และ Tree
และถ้าหาก ไม่ทำ Feature Selection จะเหมาะกับ K-Neighbors Classifier

ดังนั้นการเลือกใช้วิธีการต่างๆ ในการประมวลผลข้อมูลอย่าง Feature Selection นั้นควรพิจารณาจากลักษณะของ
ข้อมูลและวัตถุประสงค์ในการทำนายให้เหมาะสม เพื่อให้โมเดลมีประสิทธิภาพสูงสุดในการทำนายข้อมูลใหม่ที่ไม่เคยเห็นมาก่อน

ไฟล์ Data Set ก่อนและหลัง Pre-process

- Data Set ก่อน Pre-process

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

- Data Set หลัง Pre-process

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	1	89	66	23	94	28.1	0.167	21	0
1	0	137	40	35	168	43.1	2.288	33	1
2	3	78	50	32	88	31.0	0.248	26	1
3	2	197	70	45	543	30.5	0.158	53	1
4	1	189	60	23	846	30.1	0.398	59	1

บรรณานุกรม

bangkokhospital.เบาหวานกับการตั้งครรภ์. สืบค้นเมื่อ 11 มีนาคม 2567, จากเว็บไซต์:

<https://www.bangkokhospital.com/content/diabetes-and-pregnancy>

Kaggle.Diabetes Dataset. สืบค้นเมื่อ 11 มีนาคม 2567, จากเว็บไซต์:

<https://www.kaggle.com/datasets/mathchi/diabetes-data-set/data>