

## บทที่ 4

### ผลการดำเนินงาน

ผลการดำเนินงานของโครงการเรื่อง การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการเปรียบเทียบแบบจำลองที่เหมาะสมที่สุด เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

#### 4.1 ผลการวิเคราะห์ข้อมูลด้วย CRISP-DM

##### 4.1.1 ผลกระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding)

กระบวนการขั้นตอนนี้เป็นการทำความเข้าใจปัญหาและเป้าหมายเชิงธุรกิจของโครงการ รวมถึงการกำหนดขอบเขตและวัตถุประสงค์ของการนำเทคนิคการทำเหมืองข้อมูล (Data Mining) มาช่วยสนับสนุนการตัดสินใจเชิงกลยุทธ์ด้านสาธารณสุขและการวางแผนการรักษาโรคมะเร็งในอนาคต

##### 4.1.2 ผลการทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนนี้เป็นการจัดเก็บ รวบรวม และตรวจสอบความถูกต้องของข้อมูล เพื่อทำความเข้าใจคุณลักษณะของข้อมูลที่จะนำมาใช้ในการวิเคราะห์ รวมถึงการพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการดำเนินการให้สอดคล้องกับวัตถุประสงค์ของการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็ง ผู้ศึกษาทำการรวบรวมข้อมูลผู้ป่วยโรคมะเร็งจากเว็บไซต์ Kaggle (<https://www.kaggle.com>) ซึ่งเป็นแหล่งข้อมูลสาธารณะ (Open Data) ที่เผยแพร่ข้อมูลสำหรับการวิเคราะห์ทางด้านสุขภาพและชีวการแพทย์ ข้อมูลที่ใช้ประกอบด้วยประมาณ 9,062 รายการ และมี 15 ตัวแปร โดยครอบคลุมข้อมูลประชากร พฤติกรรมเสี่ยง ปัจจัยสิ่งแวดล้อม และข้อมูลทางการแพทย์ของผู้ป่วย



ภาพที่ 4.1 หน้าเว็บไซต์ kaggle.com

(ที่มา : [https://www.kaggle.com/datasets/zahidmughal2343/](https://www.kaggle.com/datasets/zahidmughal2343/global-cancer-patients-2015-2024)

[global-cancer-patients-2015-2024](https://www.kaggle.com/datasets/zahidmughal2343/global-cancer-patients-2015-2024))

#### 4.1.3 ผลการเตรียมข้อมูล (Data Preparation)

ทำการกลั่นกรองข้อมูล (Data Cleaning) โดยข้อมูลบางส่วนอาจมีความผิดพลาดหรือขาดหาย จึงจำเป็นต้องดำเนินการทำความสะอาด เช่น การลบข้อมูลที่ขาด, การเติมค่าที่หายไป, การแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสม และการตัดข้อมูลที่ไม่น่าสนใจออก เพื่อให้ข้อมูลมีความถูกต้องและสมบูรณ์สำหรับการวิเคราะห์

|    | A          | B   | C      | D         | E    | F            | G             | H           | I       | J             | K           | L            | M                  | N              | O                     |
|----|------------|-----|--------|-----------|------|--------------|---------------|-------------|---------|---------------|-------------|--------------|--------------------|----------------|-----------------------|
| 1  | Patient_ID | Age | Gender | Country_R | Year | Genetic_Risk | Air_Pollution | Alcohol_Use | Smoking | Obesity_Level | Cancer_Type | Cancer_Stage | Treatment_Cost_USD | Survival_Years | Target_Severity_Score |
| 2  | PT0000001  | 20  | Female | Brazil    | 2021 | 1.5          | 1.5           | 1.5         | 1.5     | 1.5           | 1.5 Lung    | Stage III    | 77977.12           | 2.9            | 2.2                   |
| 3  | PT0000002  | 21  | Male   | USA       | 2017 | 4.3          | 3.8           | 1           | 0.3     | 8.5           | Lung        | Stage II     | 56458.48           | 6.5            | 3.49                  |
| 4  | PT0000003  | 83  | Male   | Canada    | 2016 | 3.5          | 1.5           | 8.1         | 5       | 1.5           | Leukemia    | Stage II     | 99120.52           | 8              | 3.31                  |
| 5  | PT0000004  | 52  | Male   | Germany   | 2015 | 2.3          | 5.8           | 6.3         | 5.6     | 1.9           | Lung        | Stage II     | 19910.36           | 7              | 4.59                  |
| 6  | PT0000005  | 78  | Male   | India     | 2019 | 3.8          | 7.2           | 3.5         | 7.9     | 8.2           | Prostate    | Stage I      | 40131.04           | 5.4            | 5.77                  |
| 7  | PT0000006  | 61  | Female | India     | 2018 | 9.6          | 4.6           | 1.9         | 3.9     | 6.2           | Lung        | Stage I      | 73647.58           | 6.7            | 5.3                   |
| 8  | PT0000007  | 34  | Male   | China     | 2015 | 9.6          | 8             | 0           | 5.9     | 3.9           | Prostate    | Stage 0      | 89075.81           | 9.9            | 5.69                  |
| 9  | PT0000008  | 81  | Male   | India     | 2017 | 2.7          | 9.8           | 0.4         | 8.2     | 7.5           | Liver       | Stage I      | 12302.14           | 7.2            | 6.21                  |
| 10 | PT0000009  | 81  | Female | China     | 2018 | 0.5          | 2             | 5           | 0.7     | 2.8           | Colon       | Stage II     | 85715.44           | 6.4            | 1.86                  |
| 11 | PT0000010  | 72  | Male   | India     | 2020 | 2            | 8.8           | 8.5         | 8.1     | 6.1           | Prostate    | Stage IV     | 36662.69           | 1.8            | 6.49                  |
| 12 | PT0000011  | 21  | Female | India     | 2019 | 0.9          | 3.7           | 0.1         | 6.4     | 0.5           | Liver       | Stage I      | 40134.43           | 5.1            | 3.08                  |
| 13 | PT0000012  | 25  | Female | Germany   | 2024 | 8.4          | 6.9           | 8.5         | 2.9     | 8.1           | Breast      | Stage I      | 63821.68           | 1.5            | 5.7                   |
| 14 | PT0000013  | 23  | Male   | Canada    | 2021 | 9.6          | 8.1           | 9.8         | 2.2     | 6.9           | Prostate    | Stage II     | 17392.83           | 3.8            | 7.35                  |
| 15 | PT0000014  | 37  | Female | USA       | 2016 | 6.4          | 6.4           | 0.5         | 5.1     | 8.4           | Breast      | Stage II     | 62855.42           | 8.2            | 4.92                  |
| 16 | PT0000015  | 53  | Male   | UK        | 2018 | 6.1          | 4.4           | 4           | 7.5     | 0.1           | Skin        | Stage I      | 90874.48           | 6.5            | 4.17                  |
| 17 | PT0000016  | 81  | Male   | India     | 2016 | 4            | 1.3           | 2.6         | 6.7     | 5.2           | Prostate    | Stage II     | 92624.34           | 7.6            | 3.39                  |
| 18 | PT0000017  | 34  | Female | China     | 2020 | 7.5          | 4.5           | 7.6         | 9.2     | 5.8           | Cervical    | Stage I      | 83329.38           | 4.4            | 6.07                  |
| 19 | PT0000018  | 72  | Male   | Canada    | 2024 | 5.4          | 9.4           | 8.7         | 4.4     | 6.2           | Colon       | Stage II     | 52839.84           | 0.5            | 6.24                  |
| 20 | PT0000019  | 79  | Female | Germany   | 2022 | 8.1          | 7.4           | 6.2         | 8       | 6.4           | Cervical    | Stage III    | 47629.97           | 2.5            | 6.95                  |

ภาพที่ 4.2 (ชุดข้อมูล Global Cancer Patients 2015–2024)

#### 4.1.4 ผลการสร้างตัวแบบ (Modeling)

จากการดำเนินการสร้างโมเดลจำนวน 3 ตัวแบบ ประกอบด้วย Random Forest, Gradient Boosted Trees และ Neural Network เพื่อใช้ในการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็ง โดยใช้ข้อมูลชุด Global Cancer Patients 2015–2024 ที่รวบรวมจากเว็บไซต์ Kaggle ซึ่งมีข้อมูลประมาณ 9,062 รายการ และ 15 ตัวแปร

ในการศึกษาครั้งนี้ได้กำหนดตัวแปรอิสระ (X) เช่น อายุ ชนิดของโรคมะเร็ง ปัจจัยเสี่ยง และข้อมูลทางการแพทย์ต่าง ๆ และกำหนดตัวแปรตาม (y) คือ ระยะเวลาการอยู่รอด (Survival\_Years) จากนั้นทำการแบ่งข้อมูลออกเป็นชุดฝึกสอน (Training Set) และชุดทดสอบ (Testing Set) เพื่อป้องกันการเกิด Overfitting

โดยทำการสร้างโมเดลแต่ละประเภทโดยใช้โปรแกรม Rapidminer และทำการประเมินประสิทธิภาพด้วยตัวชี้วัดสำหรับการวิเคราะห์แบบถดถอย (Regression) ได้แก่ ค่า Root Mean Square Error (RMSE) และค่า Mean Absolute Error (MAE)

#### 4.1.5 ผลการวัดประสิทธิภาพของตัวแบบ (Evaluation)

หลังจากที่ได้ทำการสร้างแบบจำลองการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยมะเร็ง ด้วยเทคนิคทั้งสามแบบ ได้แก่ แบบจำลองโครงข่ายประสาทเทียม (Neural Network : NN) แบบจำลองแรนดอมฟอเรสต์ (Random Forest : RF) และแบบจำลองต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees : GBT) ผู้ศึกษาได้ทำการประเมินประสิทธิภาพของแต่ละแบบจำลองเพื่อเปรียบเทียบความแม่นยำของการพยากรณ์ โดยใช้ตัวชี้วัดหลักในงานวิเคราะห์เชิงพยากรณ์แบบถดถอย (Regression) ได้แก่ ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) และค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error : MAE) ซึ่งใช้ในการวัดระดับความแตกต่างระหว่างค่าที่โมเดลพยากรณ์ได้กับค่าจริง โดยค่าที่น้อยกว่าจะบ่งชี้ว่าโมเดลมีความแม่นยำสูงกว่า

##### 4.5.1 เทคนิคต้นไม้สุ่ม (Random Forest)

```

PerformanceVector

PerformanceVector:
root_mean_squared_error: 1.082 +/- 0.000
absolute_error: 0.930 +/- 0.552
  
```

ภาพที่ 4.3 ค่า RMSE และ MAE ของเทคนิคแรนดอมฟอเรสต์

#### 4.5.2 เทคนิคต้นไม้การตัดสินใจแบบไล่ระดับ (Gradient Boosted Trees)

```

PerformanceVector

PerformanceVector:
root_mean_squared_error: 1.176 +/- 0.000
absolute_error: 0.998 +/- 0.622
  
```

ภาพที่ 4.4 ค่า RMSE และ MAE ของเทคนิคต้นไม้เสริมกำลังแบบไล่ระดับ

#### 4.5.3 เทคนิคโครงข่ายประสาทเทียม (Neural Network : NN)

```

PerformanceVector

PerformanceVector:
root_mean_squared_error: 2.889 +/- 0.000
absolute_error: 2.498 +/- 1.451
  
```

ภาพที่ 4.5 ค่า RMSE และ MAE ของเทคนิคโครงข่ายประสาทเทียม

| ตัวแบบ                    | RMSE  | MAE   |
|---------------------------|-------|-------|
| Random Forest             | 1.082 | 0.938 |
| Gradient Boosted Trees    | 1.176 | 0.998 |
| Artificial Neural Network | 2.889 | 2.498 |

ตารางที่ 4.1 ตารางเปรียบเทียบประสิทธิภาพระหว่างตัวแบบ

จากตารางที่ 4.2 แสดงถึงการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบสำหรับการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการเปรียบเทียบแบบจำลองที่เหมาะสมที่สุดทั้ง 3 ตัวแบบพบว่า เทคนิค Random Forest เป็นเทคนิคที่มี ประสิทธิภาพมากที่สุด เพราะมีค่าความคลาดเคลื่อนและค่ารากของค่าคลาดเคลื่อนกำลังสอง ต่ำที่สุดเท่ากับ 1.082 และ 0.938 ในขณะที่เทคนิค Artificial Neural Network (ANN) มีค่าความคลาดเคลื่อนสูงที่สุด (RMSE = 2.889, MAE = 2.498) ซึ่งอาจเกิดจากข้อจำกัดของชุดข้อมูล หรือการปรับตั้งค่าพารามิเตอร์ของโครงข่ายประสาทเทียมที่ยังไม่เหมาะสมที่สุดสำหรับข้อมูลชุดนี้

เทคนิค Random Forest ซึ่งเป็นเทคนิคที่ชนะในข้อมูล เป็นอัลกอริทึมประเภท Ensemble Learning ที่ทำงานโดยการสร้างต้นไม้ตัดสินใจ (Decision Tree) จำนวนหลายๆ ต้นขึ้นมาเรียนรู้ข้อมูลแบบสุ่ม แล้วนำผลลัพธ์มารวมกัน โมเดลประเภทนี้จะ ไม่มีสมการเส้นตรงแบบตายตัว เหมือน Multiple Linear Regression

อย่างไรก็ตาม ในงานวิเคราะห์ข้อมูลแบบถดถอย (Regression) เพื่อทำนายตัวเลขระยะเวลา เราสามารถแสดงสมการหลักการทำงานภาพรวมของ Random Forest ได้ดังนี้

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

- $\hat{y}$  คือ ค่าพยากรณ์ระยะเวลาการอยู่รอดที่แบบจำลองทำนายได้
- B คือ จำนวนต้นไม้ตัดสินใจ (Decision Trees) ทั้งหมดที่ถูกสร้างขึ้นในโครงสร้างของ Random Forest
- $T_b(x)$  คือ ผลลัพธ์การทำนายจากต้นไม้ตัดสินใจต้นที่  $b$  โดยอิงจากชุดข้อมูลนำเข้า  $x$  (ปัจจัยต่างๆ ของผู้ป่วย)

จากสมการนี้แสดงให้เห็นว่า Random Forest ทำการพยากรณ์โดยการ หาค่าเฉลี่ย (Average) ของผลลัพธ์ที่ได้จากการทำนายของต้นไม้ตัดสินใจทุกต้นรวมกัน เพื่อลดความแปรปรวนและเพิ่มความแม่นยำ

#### 4.1.6 ผลการนำตัวแบบไปใช้งานจริง (Deployment)

ในการนำโมเดลที่ผ่านการ พัฒนาและเปรียบเทียบมาใช้งานจริง มีการดำเนินการดังต่อไปนี้

1) การเลือกตัวแบบที่เหมาะสม จากการวิเคราะห์โมเดลด้วยโปรแกรม RapidMiner ได้ทำการเปรียบเทียบค่าความ คลาดเคลื่อน (Error) ของตัวแบบทั้ง 3 ตัวแบบ พบว่าโมเดล Random forest มีค่าความคลาด เคลื่อนในการพยากรณ์อยู่ในระดับต่ำที่สุด จึงเหมาะสมที่สุด

2) สร้างโมเดล Random forest ในเว็บไซต์โดยใช้ภาษา Python และ PHP ในการ

สร้างตัวโมเดลกำหนดค่าต่างๆภายในโค้ด Python โดยตั้งค่าให้จำนวนต้นไม้ มีจำนวน 600 และความลึกสูงสุด 50 และกำหนดปัจจัยในการพยากรณ์โดยเลือก ปีที่มีชีวิตรอด ใช้เป็น Label อายุ ความเสี่ยงทางพันธุกรรม ความเสี่ยงมลพิษทางอากาศ ปริมาณการดื่ม ปริมาณการสูบบุหรี่ ค่าความอ้วน ระยะโรคมะเร็ง ปีที่ค้นพบโรคมะเร็งให้ตรงกับโมเดล Random Forest ภายใน Rapidminer

3) ทำการเทรน Random forest โมเดลโดยใช้ข้อมูล CSV ในการเทรน และสร้างเป็นไฟล์โมเดลสำหรับการพยากรณ์ขึ้นมาและนำไปใส่ในตัวเว็บไซต์เพื่อใช้ในการพยากรณ์

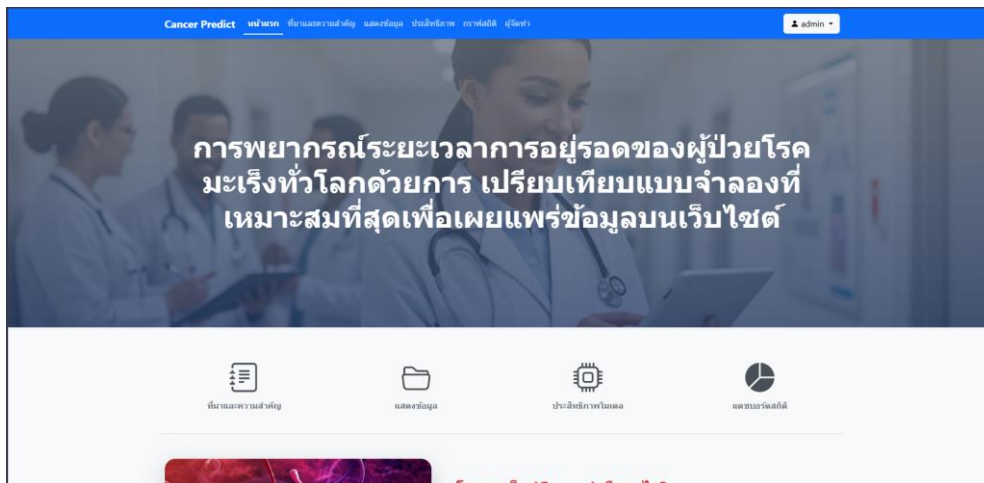
4) การนำเสนอผ่านเว็บไซต์ นำโมเดล Random forest มาทำการพยากรณ์ได้โดยการใส่ข้อมูลเข้าไปเพื่อพยากรณ์ ทันที่ (Random Forest บนเว็บไซต์ โดยใช้ PHP Java Script และ Python)

## 4.2 ผลการออกแบบเว็บไซต์

4.2.1 ผลการวิเคราะห์และออกแบบฐานข้อมูลผลการวิเคราะห์และออกแบบฐานข้อมูลระบบฐานข้อมูลนี้ถูกออกแบบมาเพื่อจัดเก็บข้อมูลเพื่อใช้ในการล็อกอินเข้าระบบ โดยโครงสร้างของฐานข้อมูลนี้ประกอบด้วยหลายฟิลด์หลัก เช่น ชื่อผู้ใช้ (Username ) และ รหัสผ่าน (Password) ระบบยังมีสิทธิ์การใช้งานของระบบแบ่งออกเป็น 2 ระดับหลัก ได้แก่ ผู้ดูแลระบบ (Admin) และผู้ใช้งานทั่วไป (User) ผู้ใช้งานทั่วไปสามารถล็อกอิน เข้าสู่ระบบและเข้าถึงข้อมูลที่ได้รับอนุญาตเท่านั้น สามารถแก้ไขหรือลบข้อมูลได้ สามารถ ค้นหาข้อมูล ดาวน์โหลดข้อมูลในรูปแบบ CSV ได้เพื่อนำไปต่อยอดได้ ซึ่งแสดงให้เห็นว่า ฐานข้อมูลถูกออกแบบให้รองรับการบริหารจัดการข้อมูลขนาดใหญ่ และสามารถดึงข้อมูลออกไป วิเคราะห์เพิ่มเติมในเครื่องมือภายนอกได้

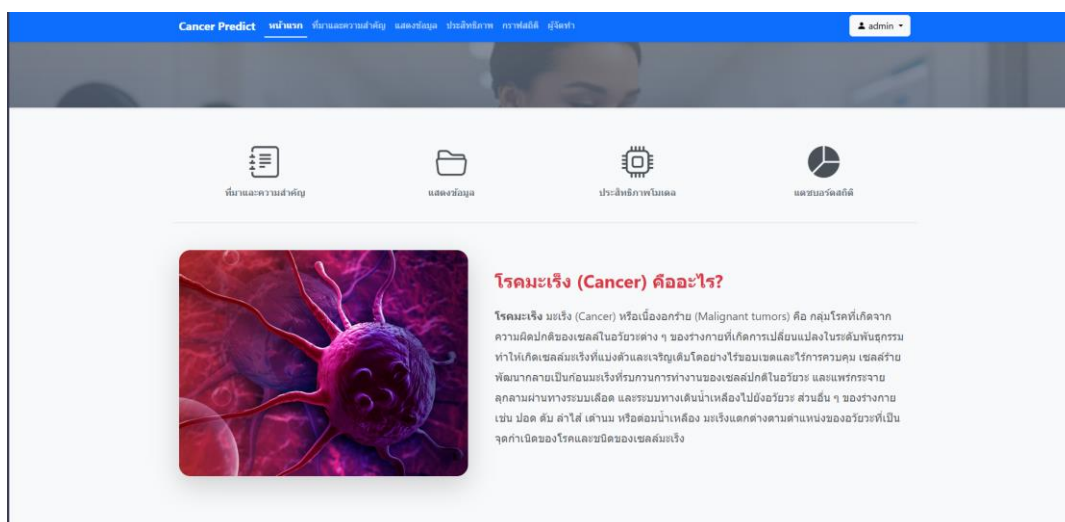
4.2.2 ผลการออกแบบเว็บไซต์การออกแบบหน้าแรกของเว็บไซต์นำเสนอเมนูต่าง ๆ และเนื้อหา ที่เกี่ยวข้องกับการเปรียบเทียบโมเดลเพื่อพยากรณ์ระยะเวลาผู้ป่วยโรคมะเร็งทั่วโลก โดยการแสดงผลข้อมูลที่ผ่านมาการพยากรณ์ ข้อมูลสารสนเทศนี้ถูกเผยแพร่บนเว็บเบราว์เซอร์ เพื่อให้ผู้ใช้งานทั่วไปสามารถเข้าถึงและเข้าใจข้อมูลได้ อย่างสะดวกรวดเร็วและถูกต้อง ผู้

วิเคราะห์ข้อมูลได้ดำเนินการวิเคราะห์และออกแบบเว็บไซต์ตาม วัตถุประสงค์ที่กำหนดไว้ดังต่อไปนี้



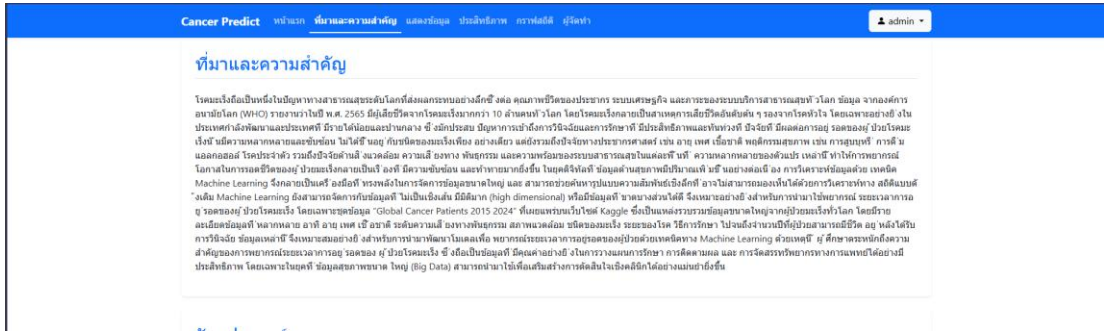
ภาพที่ 4.6 แสดงหน้าหลักของเว็บไซต์

จากภาพที่ 4.6 หน้าเว็บเพจ การเปรียบเทียบโมเดลเพื่อพยากรณ์ระยะเวลาผู้ป่วยโรคมะเร็งทั่วโลก โดยมีส่วนหัว (Header) ที่ประกอบด้วยเมนูนำทาง (Navigation Menu) และข้อความหัวข้อหลัก (Main Title) แถบเมนูสามารถใช้งานได้ตามความต้องการของการใช้งาน เป็นหน้าสำหรับผู้ใช้งานทั่วไปสามารถเลือกดูและศึกษาค้นคว้าเข้าไปใช้งานในส่วนของเว็บไซต์



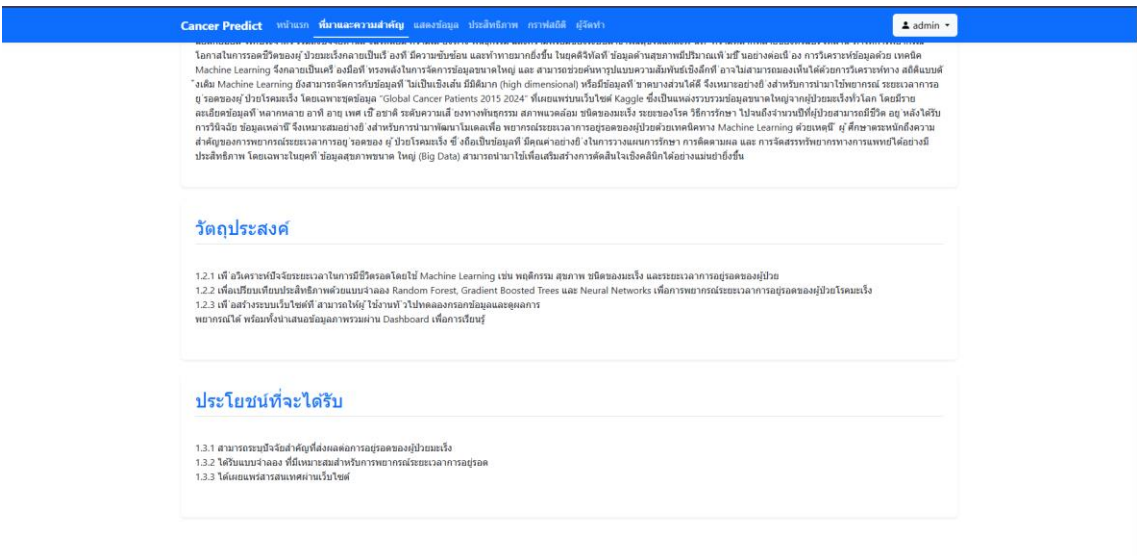
ภาพที่ 4.7 แสดงข้อมูลเพิ่มเติมของหน้าหลักเว็บไซต์

จากภาพที่ 4.7 หน้าเว็บจะแสดงให้เห็นเมนูการนำทางไปหน้าต่างๆโดยจะมี ที่มาและความสำคัญ แสดงข้อมูลผู้ป่วย ประสิทธิภาพของโมเดล และแดชบอร์ดสถิติ เป็นตัวเลือก สำหรับผู้ใช้งานทั่วไปสามารถเลือกดูได้ และยังมีการอธิบายความรู้เกี่ยวกับโรคมะเร็ง



ภาพที่ 4.8 แสดงหน้าที่มาและความสำคัญ

จากภาพที่ 4.8 เป็นหน้าที่แสดงการอธิบายที่มาและความสำคัญของการเปรียบเทียบโมเดลเพื่อเพื่อพยากรณ์ระยะเวลาผู้ป่วยโรคมะเร็งทั่วโลก



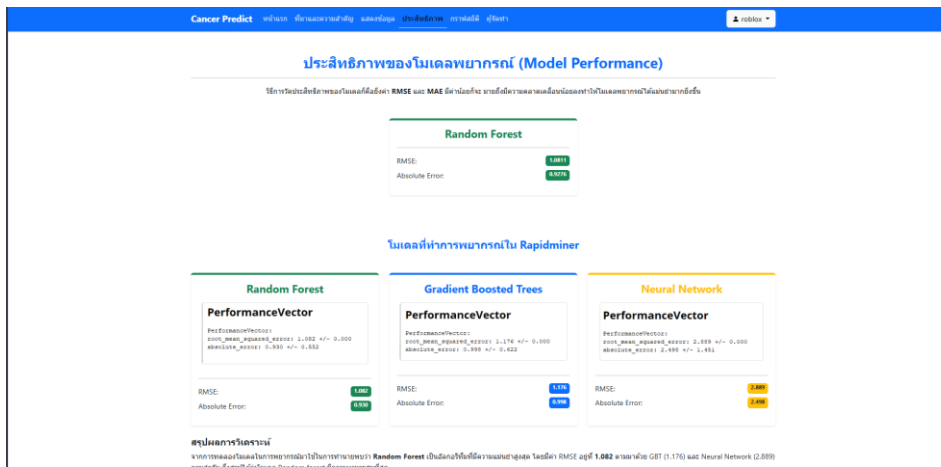
ภาพที่ 4.9 แสดงหน้าเพิ่มเติมของที่มาและความสำคัญ

จากภาพที่ 4.9 เป็นหน้าที่แสดงการอธิบายวัตถุประสงค์และประโยชน์ของการเปรียบเทียบโมเดลเพื่อเพื่อพยากรณ์ระยะเวลาผู้ป่วยโรคมะเร็งทั่วโลก

| ผู้ป่วย  | เพศ    | สัญชาติ | ปีเกิด | โรคมะเร็งที่พยากรณ์ | โรคมะเร็งที่แท้จริง | ระยะเวลาอยู่รอด |
|----------|--------|---------|--------|---------------------|---------------------|-----------------|
| P1000001 | Female | Spain   | 2021   | ESIT (Breast)       | ESIT (Breast)       | 4 ปี 9 เดือน    |
| P1000002 | Male   | USA     | 2017   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 9 เดือน    |
| P1000003 | Male   | Canada  | 2016   | ESIT (Lung)         | ESIT (Lung)         | 7 ปี 2 เดือน    |
| P1000004 | Male   | Germany | 2015   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 10 เดือน   |
| P1000005 | Male   | India   | 2019   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 2 เดือน    |
| P1000006 | Female | India   | 2018   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 8 เดือน    |
| P1000007 | Male   | China   | 2015   | ESIT (Lung)         | ESIT (Lung)         | 6 ปี 3 เดือน    |
| P1000008 | Male   | India   | 2017   | ESIT (Lung)         | ESIT (Lung)         | 4 ปี 1 เดือน    |
| P1000009 | Female | China   | 2018   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 9 เดือน    |
| P1000010 | Male   | India   | 2020   | ESIT (Lung)         | ESIT (Lung)         | 3 ปี            |
| P1000011 | Female | India   | 2019   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 4 เดือน    |
| P1000012 | Female | Germany | 2024   | ESIT (Lung)         | ESIT (Lung)         | 2 ปี 8 เดือน    |
| P1000013 | Male   | Canada  | 2021   | ESIT (Lung)         | ESIT (Lung)         | 4 ปี 3 เดือน    |
| P1000014 | Female | USA     | 2018   | ESIT (Lung)         | ESIT (Lung)         | 7 ปี            |
| P1000015 | Male   | UK      | 2018   | ESIT (Lung)         | ESIT (Lung)         | 6 ปี 2 เดือน    |
| P1000016 | Male   | India   | 2016   | ESIT (Lung)         | ESIT (Lung)         | 6 ปี 1 เดือน    |
| P1000017 | Female | China   | 2020   | ESIT (Lung)         | ESIT (Lung)         | 4 ปี 5 เดือน    |
| P1000018 | Male   | Canada  | 2024   | ESIT (Lung)         | ESIT (Lung)         | 2 ปี 11 เดือน   |
| P1000019 | Female | Canada  | 2021   | ESIT (Lung)         | ESIT (Lung)         | 5 ปี 6 เดือน    |

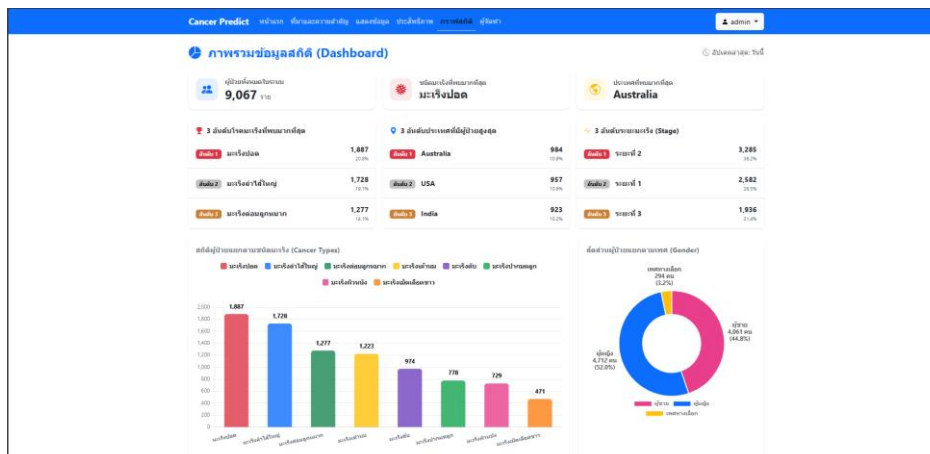
ภาพที่ 4.10 หน้าแสดงข้อมูล

จากภาพที่ 4.10 เป็นหน้าที่แสดงข้อมูลของผู้ป่วย ค่าต่างๆที่เกี่ยวกับโรคมะเร็ง และแสดงผลพยากรณ์ ระยะเวลาอยู่รอด



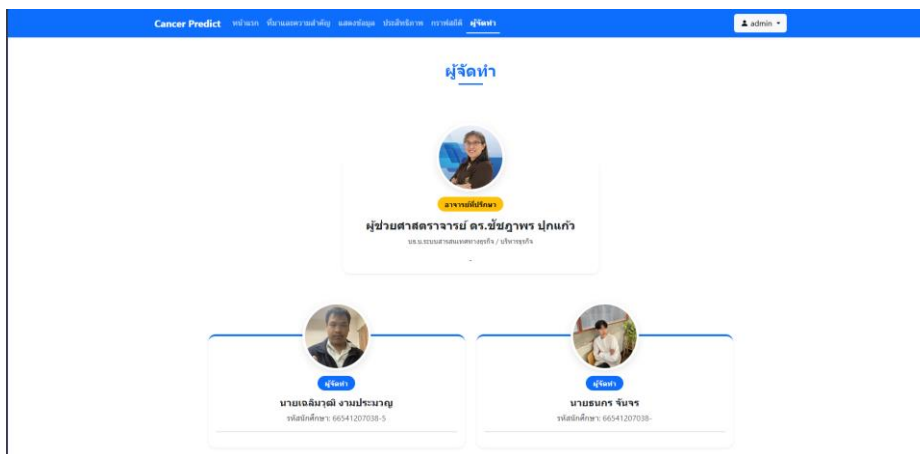
ภาพที่ 4.11 หน้าแสดงประสิทธิภาพของโมเดล

จากภาพที่ 4.11 เป็นหน้าที่แสดงค่าความแม่นยำของโมเดลที่ใช้ภายในเว็บไซต์และมีส่วนแสดงการเปรียบเทียบโมเดล 3 ทั้งนี้ นำมาจาก RapidMiner และสร้างขึ้นมาในเว็บไซต์



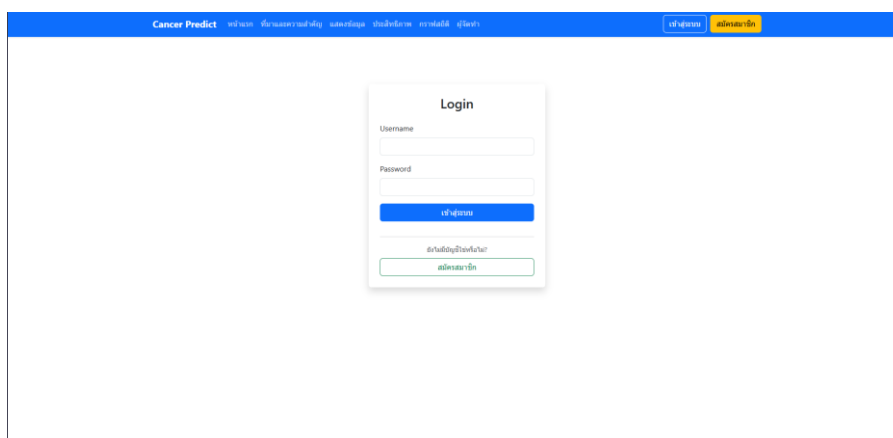
ภาพที่ 4.12 หน้าแสดงแดชบอร์ด

จากภาพที่ 4.12 เป็นหน้าที่แสดงผลแดชบอร์ดของข้อมูลภายในเว็บไซต์



ภาพที่ 4.13 หน้าแสดงผู้จัดทำ

จากภาพที่ 4.13 เป็นหน้าที่ผู้ที่จัดเว็บไซต์ และผู้ปรึกษาโครงการ



ภาพที่ 4.14 หน้าแสดงเข้าสู่ระบบ

จากภาพที่ 4.14 เป็นหน้าที่ใช้ในการเข้าสู่ระบบหรือสมัครสมาชิก

Cancer Predict หน้าหลัก ติดต่อเราเกี่ยวกับศูนย์ บริการข้อมูล บริการปรึกษา การพบแพทย์ ผู้พัฒนา

เข้าสู่ระบบ สมัครสมาชิก

### สมัครสมาชิก

**Username \***  
 👤 ชื่อผู้ใช้ระบบ

**Email \***  
 ✉ ชื่อระบบ: user@example.com

**Password \***  
 🔑 รหัสผ่าน

ยืนยันการสมัครสมาชิก

ภาพที่ 4.15 หน้าสมัครสมาชิก

จากภาพที่ 4.15 เป็นหน้าที่ใช้ในการสมัครสมาชิกเพื่อใช้งานเว็บไซต์

Cancer Predict หน้าหลัก ติดต่อเราเกี่ยวกับศูนย์ บริการข้อมูล บริการปรึกษา การพบแพทย์ ผู้พัฒนา

admin

### แก้ไขโปรไฟล์

**Username (ไม่สามารถแก้ไขได้)**  
 admin

**Email**  
 ✉ admin@admin.com

**รหัสผ่าน**  
 🔑 |

บันทึกการเปลี่ยนแปลง

ยกเลิก

ภาพที่ 4.16 หน้าแก้ไขโปรไฟล์

จากภาพที่ 4.16 เป็นหน้าที่ใช้ในการแก้ไขรหัสผ่าน และอีเมล

Cancer Predict หน้าแรก | บริการทางการแพทย์ | แสดงข้อมูล | ประวัติการพยากรณ์ | เข้าสู่ระบบ admin

**ประวัติการพยากรณ์**  
ประวัติการพยากรณ์

ค้นหาประวัติการพยากรณ์...

| รหัสผู้ป่วย | วันที่พยากรณ์    | ระยะเวลาของผล | ผลพยากรณ์     | ผู้พยากรณ์ | ผลการพยากรณ์ |
|-------------|------------------|---------------|---------------|------------|--------------|
| PT0000001   | 2020-03-17 08:28 | 0.11 เดือน    | 4 ปี 9 เดือน  | admin      | ✓ ✗          |
| PT0000002   | 2020-03-17 08:28 | 0.13 เดือน    | 5 ปี 9 เดือน  | admin      | ✓ ✗          |
| PT0000003   | 2020-03-17 08:28 | 0.3           | 7 ปี 2 เดือน  | admin      | ✓ ✗          |
| PT0000004   | 2020-03-17 08:28 | 0.3           | 5 ปี 10 เดือน | admin      | ✓ ✗          |
| PT0000005   | 2020-03-17 08:28 | 0.33 เดือน    | 5 ปี 2 เดือน  | admin      | ✓ ✗          |
| PT0000006   | 2020-03-17 08:28 | 0.33 เดือน    | 5 ปี 8 เดือน  | admin      | ✓ ✗          |
| PT0000007   | 2020-03-17 08:28 | 0.11 เดือน    | 6 ปี 3 เดือน  | admin      | ✓ ✗          |
| PT0000008   | 2020-03-17 08:28 | 0.22 เดือน    | 6 ปี 1 เดือน  | admin      | ✓ ✗          |
| PT0000009   | 2020-03-17 08:28 | 0.33 เดือน    | 5 ปี 9 เดือน  | admin      | ✓ ✗          |
| PT0000010   | 2020-03-17 08:28 | 1.01 เดือน    | 3 ปี          | admin      | ✓ ✗          |
| PT0000011   | 2020-03-17 08:28 | 0.11 เดือน    | 5 ปี 4 เดือน  | admin      | ✓ ✗          |
| PT0000012   | 2020-03-17 08:28 | 1.01 เดือน    | 2 ปี 8 เดือน  | admin      | ✓ ✗          |
| PT0000013   | 2020-03-17 08:28 | 0.19 เดือน    | 4 ปี 3 เดือน  | admin      | ✓ ✗          |
| PT0000014   | 2020-03-17 08:28 | 0.11 เดือน    | 7 ปี          | admin      | ✓ ✗          |
| PT0000015   | 2020-03-17 08:28 | 0.33 เดือน    | 6 ปี 2 เดือน  | admin      | ✓ ✗          |
| PT0000016   | 2020-03-17 08:28 | 0.33 เดือน    | 6 ปี 1 เดือน  | admin      | ✓ ✗          |

ภาพที่ 4.17 หน้าประวัติการพยากรณ์

จากภาพที่ 4.17 เป็นหน้าที่ใช้ในการดูประวัติการพยากรณ์ที่ผู้ใช้เคยพยากรณ์

#### 4.4 บทสรุป

การศึกษานี้ได้ทำการหาตัวแบบที่เหมาะสม เพื่อพยากรณ์ระยะเวลารอดชีวิตของผู้ป่วยโรคมะเร็ง โดยพิจารณาปัจจัยเสี่ยงทางพฤติกรรมและข้อมูลทางการแพทย์ที่มีผลกระทบต่ออัตราการอยู่รอด รวมถึงการจัดทำระบบเว็บแอปพลิเคชันสำหรับการจัดการและแสดงผลข้อมูลอย่างเป็นรูปธรรม

จากผลการวิเคราะห์และพัฒนาแบบจำลอง พบว่าอัลกอริทึม Random Forest สามารถนำมาประยุกต์ใช้ในการพยากรณ์ระยะเวลารอดชีวิตได้อย่างมีประสิทธิภาพ โดยอาศัยตัวแปรสำคัญที่มีอิทธิพลต่อการพยากรณ์ ได้แก่ ข้อมูลประชากรศาสตร์ (อายุ, เพศ, ประเทศ) ปัจจัยเสี่ยงทางพฤติกรรมและสิ่งแวดล้อม (พันธุกรรม, มลพิษ, การดื่มสุรา, การสูบบุหรี่, ภาวะโรคอ้วน) และข้อมูลทางการแพทย์ (ชนิดของมะเร็ง, ระยะของมะเร็ง, ระดับความรุนแรง) ทั้งนี้ ผู้ศึกษาได้ออกแบบระบบโดยข้อมูลใช้จาก Kaggle เพื่อใช้ในการทดสอบประสิทธิภาพการทำงานของโมเดลและเพื่อหาโมเดลที่แม่นยำที่สุดเพื่อมาใช้ในการพยากรณ์ระยะเวลาอยู่รอด ซึ่งก็คือโมเดล Random Forest ที่มีค่า RMSE = 1.0738 ปี และ MAE = 0.9244

นอกจากนี้ส่วนของระบบเว็บไซต์ที่พัฒนาขึ้น ผู้ใช้งานสามารถเข้าถึงหน้าแดชบอร์ดสถิติ (Dashboard) เพื่อดูภาพรวมและแนวโน้มอัตราการอยู่รอดเฉลี่ย ซึ่งแสดงผลในรูปแบบกราฟที่สามารถจำแนกตามกลุ่มอายุ เพศ และประเทศที่พบ นอกจากนี้ ระบบยังมีหน้าต่างสำหรับการจัดการข้อมูล (Data Management) ที่อนุญาตให้ผู้ใช้งานเพิ่ม แก้ไข ลบ และสั่งรันการพยากรณ์ผลผู้ป่วยเป็นรายบุคคลได้แบบเรียลไทม์ ในขณะที่ผู้ดูแลระบบ (Admin) สามารถจัดการข้อมูลในภาพรวมด้วยฟังก์ชันการนำเข้าและส่งออกชุดข้อมูลผ่านไฟล์ CSV ได้อย่างสะดวกและรวดเร็ว

ซึ่งการศึกษานี้ได้นำเสนอโครงสร้างระบบการพยากรณ์ทางการแพทย์ ตั้งแต่การบูรณาการแบบโมเดลพยากรณ์ เข้ากับระบบฐานข้อมูล ไปจนถึงการพัฒนาเครื่องมือบนเว็บไซต์ที่ช่วยให้สามารถนำข้อมูลไปใช้ประโยชน์ได้ง่ายขึ้น ซึ่งระบบดังกล่าวสามารถนำไปพัฒนาต่อยอดเพื่อรองรับการวิเคราะห์ข้อมูลผู้ป่วยจริง ช่วยสนับสนุนการตัดสินใจทางการแพทย์ และการวางแผนสาธารณสุขในอนาคตได้อย่างมีประสิทธิภาพ