บทที่ 3
ขั้นตอน และ วิธีการดำเนินงาน

โครงงานเรื่อง การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการเปรียบเทียบแบบจำลองที่เหมาะสมที่สุด เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้จะมีขั้นตอนและรายละเอียดดังนี้ โดยทำการเปรียบเทียบตัวแบบ 3 ตัวแบบ คือRandom Forest (RF)โมเดลแบบ Ensemble ที่สร้างDecision Tree หลาย ๆ ต้นมารวมผลลัพธ์เข้าด้วยกันพยากรณ์ Gradient Boosted Trees (GBDT / GBM)โมเดลที่ปรับค่าพารามิเตอร์อย่างต่อเนื่องทีละขั้น เพื่อให้ผลลัพธ์ใกล้เคียงค่าจริงมากขึ้น Neural Networks (NNs / Deep Learning)แบบจำลองที่เลียนแบบโครงสร้างสมองมนุษย์ ใช้ Hidden Layers หลายชั้นในการเรียนรู้และนำมาเปรียบเทียบประสิทธิภาพของตัวแบบ โดยใช้ค่า RMSE และ MAE เป็นตัวเปรียบเทียบ จากนั้นทำการนำตัวแบบที่ได้ไปทำการพัฒนาเว็บไซต์ สำหรับการแสดงผล

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

กระบวนการวิเคราะห์ข้อมูลด้วย Cross-Industry Standard Process for Data Mining หรือ CRISP-DM พัฒนาขึ้นปีในปี ค.ศ. 1996 เป็นขั้นตอนการทำเหมืองข้อมูลที่ได้รับการยอมรับใน ระดับ สากล ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding)

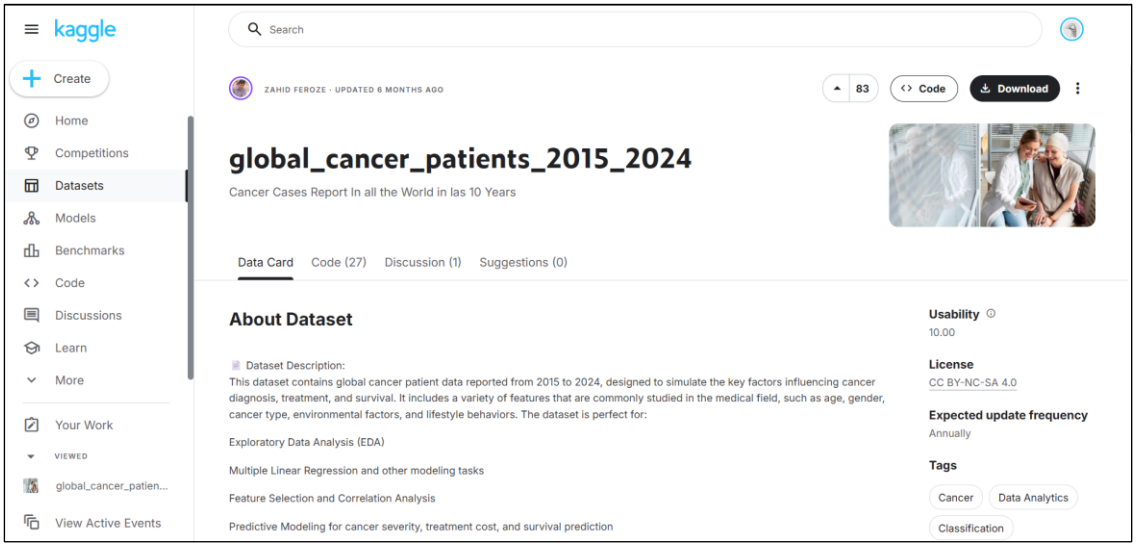
ขั้นตอนนี้เป็นการทำความเข้าใจปัญหาและเป้าหมายเชิงธุรกิจของโครงงาน รวมถึงการกำหนดขอบเขตและวัตถุประสงค์ของการนำเทคนิคการทำเหมืองข้อมูล (Data Mining) มาช่วยสนับสนุนการตัดสินใจเชิงกลยุทธ์ด้านสาธารณสุขและการวางแผนการรักษาโรคมะเร็งในอนาคต

3.1.2 ทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนนี้เป็นการจัดเก็บ รวบรวม และตรวจสอบความถูกต้องของข้อมูล เพื่อทำความเข้าใจคุณลักษณะของข้อมูลที่นำมาใช้ในการวิเคราะห์ รวมถึงการพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือ

บางส่วนการดำเนินการให้สอดคล้องกับวัตถุประสงค์ของการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็ง

ผู้ศึกษาทำการรวบรวมข้อมูลผู้ป่วยโรคมะเร็งจากเว็บไซต์ **Kaggle** (https://www.kaggle.com) ซึ่งเป็นแหล่งข้อมูลสาธารณะ (Open Data) ที่เผยแพร่ข้อมูลสำหรับการวิเคราะห์ทางด้านสุขภาพและชีวการแพทย์ ข้อมูลที่ใช้ประกอบด้วยประมาณ **50,000** รายการ และมี 15 ตัวแปร โดยครอบคลุมข้อมูลประชากร พฤติกรรมเสี่ยง ปัจจัยสิ่งแวดล้อม และข้อมูลทางการแพทย์ของผู้ป่วย

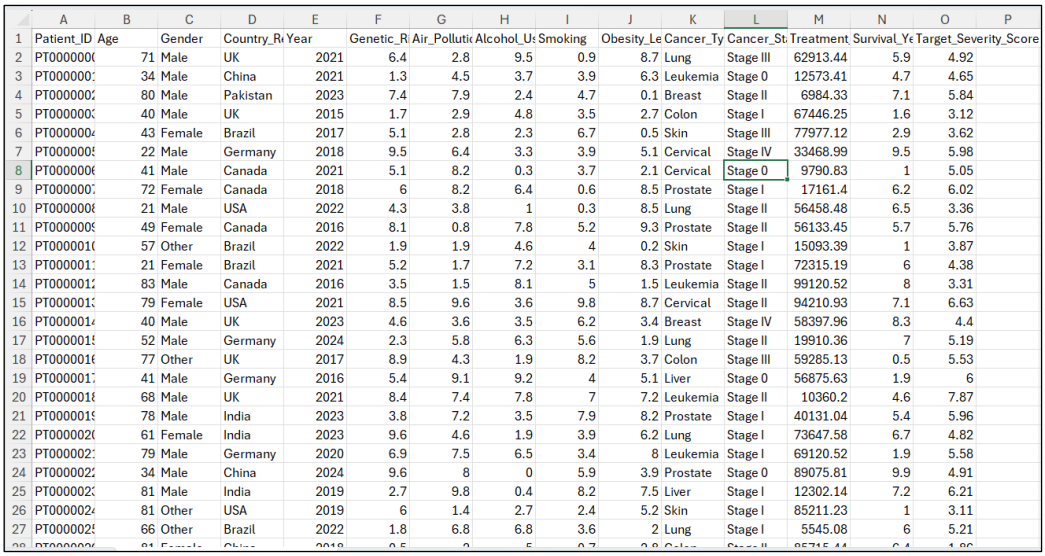


ภาพที่ 3.1 หน้าเว็บไซต์ kaggle.com

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

3.1.3 การทำเตรียมข้อมูล (Data Preparation)

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



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Patient_ID	Age	Gender	Country_R	Year	Genetic_R	Air_Pollutio	Alcohol_Us	Smoking	Obesity_Le	Cancer_Ty	Cancer_St	Treatment	Survival_Ye	Target_Severity_Score	
2	PT000000(	71	Male	UK	2021	6.4	2.8	9.5	0.9	8.7	Lung	Stage III	62913.44	5.9	4.92	
3	PT000000	34	Male	China	2021	1.3	4.5	3.7	3.9	6.3	Leukemia	Stage 0	12573.41	4.7	4.65	
4	PT000000	80	Male	Pakistan	2023	7.4	7.9	2.4	4.7	0.1	Breast	Stage II	6984.33	7.1	5.84	
5	PT000000	40	Male	UK	2015	1.7	2.9	4.8	3.5	2.7	Colon	Stage I	67446.25	1.6	3.12	
6	PT000000	43	Female	Brazil	2017	5.1	2.8	2.3	6.7	0.5	Skin	Stage III	77977.12	2.9	3.62	
7	PT000000	22	Male	Germany	2018	9.5	6.4	3.3	3.9	5.1	Cervical	Stage IV	33468.99	9.5	5.98	
8	PT000000	41	Male	Canada	2021	5.1	8.2	0.3	3.7	2.1	Cervical	Stage 0	9790.83	1	5.05	
9	PT000000	72	Female	Canada	2018	6	8.2	6.4	0.6	8.5	Prostate	Stage I	17161.4	6.2	6.02	
10	PT000000	21	Male	USA	2022	4.3	3.8	1	0.3	8.5	Lung	Stage II	56458.48	6.5	3.36	
11	PT000000	49	Female	Canada	2016	8.1	0.8	7.8	5.2	9.3	Prostate	Stage II	56133.45	5.7	5.76	
12	PT000001(	57	Other	Brazil	2022	1.9	1.9	4.6	4	0.2	Skin	Stage I	15093.39	1	3.87	
13	PT000001	21	Female	Brazil	2021	5.2	1.7	7.2	3.1	8.3	Prostate	Stage I	72315.19	6	4.38	
14	PT000001	83	Male	Canada	2016	3.5	1.5	8.1	5	1.5	Leukemia	Stage II	99120.52	8	3.31	
15	PT000001	79	Female	USA	2021	8.5	9.6	3.6	9.8	8.7	Cervical	Stage II	94210.93	7.1	6.63	
16	PT000001	40	Male	UK	2023	4.6	3.6	3.5	6.2	3.4	Breast	Stage IV	58397.96	8.3	4.4	
17	PT000001	52	Male	Germany	2024	2.3	5.8	6.3	5.6	1.9	Lung	Stage II	19910.36	7	5.19	
18	PT000001	77	Other	UK	2017	8.9	4.3	1.9	8.2	3.7	Colon	Stage III	59285.13	0.5	5.53	
19	PT000001	41	Male	Germany	2016	5.4	9.1	9.2	4	5.1	Liver	Stage 0	56875.63	1.9	6	
20	PT000001	68	Male	UK	2021	8.4	7.4	7.8	7	7.2	Leukemia	Stage II	10360.2	4.6	7.87	
21	PT000001	78	Male	India	2023	3.8	7.2	3.5	7.9	8.2	Prostate	Stage I	40131.04	5.4	5.96	
22	PT000002(	61	Female	India	2023	9.6	4.6	1.9	3.9	6.2	Lung	Stage I	73647.58	6.7	4.82	
23	PT000002	79	Male	Germany	2020	6.9	7.5	6.5	3.4	8	Leukemia	Stage I	69120.52	1.9	5.58	
24	PT000002	34	Male	China	2024	9.6	8	0	5.9	3.9	Prostate	Stage 0	89075.81	9.9	4.91	
25	PT000002	81	Male	India	2019	2.7	9.8	0.4	8.2	7.5	Liver	Stage I	12302.14	7.2	6.21	
26	PT000002	81	Other	USA	2019	6	1.4	2.7	2.4	5.2	Skin	Stage I	85211.23	1	3.11	
27	PT000002	66	Other	Brazil	2022	1.8	6.8	6.8	3.6	2	Lung	Stage I	5545.08	6	5.21	

ภาพที่ 3.2 (ชุดข้อมูล *Global Cancer Patients 2015–2024*)

3.1.4 การสร้างตัวแบบ (Modeling)

3.1.4.1 การสร้างตัวแบบการพยากรณ์ โดยใช้เทคนิคต้นไม้สุ่ม (Random Forest) การสร้างตัวแบบการพยากรณ์ ด้วยเทคนิค Random Forest จะดำเนินการผ่านโปรแกรม RapidMiner โดยมีขั้นตอนดังนี้

1. นำเข้าชุดข้อมูลที่ผ่านการเตรียมความพร้อมแล้ว (Retrieve Data)
2. เลือกปัจจัยที่จำเป็นสำหรับการพยากรณ์ (Select Attribute)
3. แปลงข้อมูลบางส่วนเพื่อที่จะนำไปใช้ในโมเดล
4. กำหนดบทบาทของตัวแปร (Set Role) โดยให้ตัวแปรระยะเวลารอดชีวิต (Survival Years) เป็นตัวแปรเป้าหมาย (Label)
5. นำข้อมูลส่วนฝึกสอนเข้าสู่โอเปอเรเตอร์ Random Forest เพื่อให้ระบบเรียนรู้และสร้างแบบจำลอง
6. นำแบบจำลองที่ได้ไปเชื่อมต่อกับโอเปอเรเตอร์ Apply Model ร่วมกับข้อมูลส่วนทดสอบ เพื่อพยากรณ์ผลลัพธ์
7. ประเมินความแม่นยำและหาค่าความคลาดเคลื่อนด้วยโอเปอเรเตอร์Performance (Regression)โดยใช้ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยราก(RMSE)

3.1.4.2 การสร้างตัวแบบการพยากรณ์ โดยใช้เทคนิคต้นไม้การตัดสินใจแบบไล่ระดับ (Gradient Boosted Trees) การสร้างตัวแบบการพยากรณ์ ด้วยเทคนิค Gradient Boosted Trees จะดำเนินการผ่านโปรแกรม RapidMiner โดยมีขั้นตอนดังนี้

1. นำเข้าชุดข้อมูลที่ผ่านการเตรียมความพร้อมแล้ว (Retrieve Data)

2. เลือกปัจจัยที่จำเป็นสำหรับการพยากรณ์ (Select Attribute)

3. แปลงข้อมูลบางส่วนเพื่อที่จะนำไปใช้ในโมเดล

4. กำหนดบทบาทของตัวแปร (Set Role) โดยให้ตัวแปรระยะเวลารอดชีวิต (Survival Years) เป็นตัวแปรเป้าหมาย (Label)

5. นำข้อมูลเข้าสู่โอเปอเรเตอร์ Gradient Boosted Trees ซึ่งอัลกอริทึมจะทำการสร้างต้นไม้การตัดสินใจและปรับปรุงค่าความผิดพลาด (Error) อย่างต่อเนื่องทีละขั้นตอน

6. นำแบบจำลองที่ได้ไปเชื่อมต่อกับโอเปอเรเตอร์ Apply Model ร่วมกับข้อมูลส่วนทดสอบ เพื่อพยากรณ์ผลลัพธ์

7. ประเมินความแม่นยำและหาค่าความคลาดเคลื่อนด้วยโอเปอเรเตอร์ Performance (Regression) โดยใช้ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยราก (RMSE)

3.1.4.3 การสร้างตัวแบบการพยากรณ์ โดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network : NN) การสร้างตัวแบบการพยากรณ์ ด้วยเทคนิค Neural Network จะดำเนินการผ่านโปรแกรม RapidMiner โดยมีขั้นตอนดังนี้

1. นำเข้าชุดข้อมูลที่ผ่านการเตรียมความพร้อมแล้ว (Retrieve Data)

2. เลือกปัจจัยที่จำเป็นสำหรับการพยากรณ์ (Select Attribute)

3. แปลงข้อมูลบางส่วนเพื่อที่จะนำไปใช้ในโมเดล

4. กำหนดบทบาทของตัวแปร (Set Role) โดยให้ตัวแปรระยะเวลารอดชีวิต (Survival Years) เป็นตัวแปรเป้าหมาย (Label)

5. นำชุดข้อมูลที่ผ่านการเตรียมความพร้อมเข้าสู่โอเปอเรเตอร์ Neural

Networkซึ่งจำลองการทำงานของโครงข่ายประสาทในสมองมนุษย์ กำหนดจำนวนชั้นซ่อน (Hidden Layers) และรอบการเรียนรู้ (Training Cycles) เพื่อให้แบบจำลองสามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนของปัจจัยเสี่ยงต่าง ๆ

6. ทดสอบการพยากรณ์ (Apply Model): นำโมเดลที่ฝึกสอนเสร็จแล้วไปใช้พยากรณ์ (Predict) ค่าจากชุดข้อมูลทดสอบ

7. แสดงผลการวัดประสิทธิภาพ (Performance) วัดค่าความแม่นยำในการพยากรณ์ด้วยตัวชี้วัดทางสถิติ (RMSE, MAE,)เพื่อดูความคลาดเคลื่อนของตัวเลขระยะเวลาที่พยากรณ์ ได้

3.1.4.4 การสร้างและปรับใช้แบบจำลองสำหรับการพยากรณ์ด้วยภาษา Python หลังจากได้ผลสรุปว่าโมเดลใดมีประสิทธิภาพดีที่สุดจากโปรแกรม RapidMiner ผู้ศึกษาได้นำอัลกอริทึมที่เหมาะสมที่สุดมาดำเนินการต่อ ดังนี้

1. นำอัลกอริทึม Random Forest Regressor มาพัฒนาเชิงลึก (Implementation) โดยใช้ภาษา Python ร่วมกับไลบรารีสำหรับการเรียนรู้ของเครื่อง (scikit-learn)

2. ทำการเขียนโปรแกรมสำหรับฝึกสอนแบบจำลองอีกครั้งและบันทึกแบบจำลองนั้นเป็นไฟล์ (เช่น .pkl หรือ .joblib)

3. เชื่อมต่อแบบจำลองเข้ากับระบบเว็บไซต์ (Web Application)

4. ระบบที่พัฒนาด้วย Python จะทำหน้าที่รับค่าพารามิเตอร์ปัจจัยเสี่ยงของผู้ป่วยที่ส่งมาจากหน้าเว็บไซต์ ทำการประมวลผลการพยากรณ์ และส่งค่าผลลัพธ์ (จำนวนปีที่คาดว่าจะรอดชีวิต) กลับไปแสดงผลบนหน้าเว็บไซต์ได้แบบเรียลไทม์

3.1.5 การวัดประสิทธิภาพของโมเดล

หลังจากที่ได้ทำการสร้างแบบจำลองการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยมะเร็งด้วยเทคนิคทั้งสามแบบ ได้แก่ แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Network : ANN) 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

ตารางที่ 3.1 แสดงข้อมูลเปรียบเทียบค่าความแม่นยำของโมเดลพยากรณ์

เทคนิค	RMSE	MAE
Random Forest	1.082	0.930
Gradient Boosted Tree	1.176	0.998
Neural Network	2.889	2.498

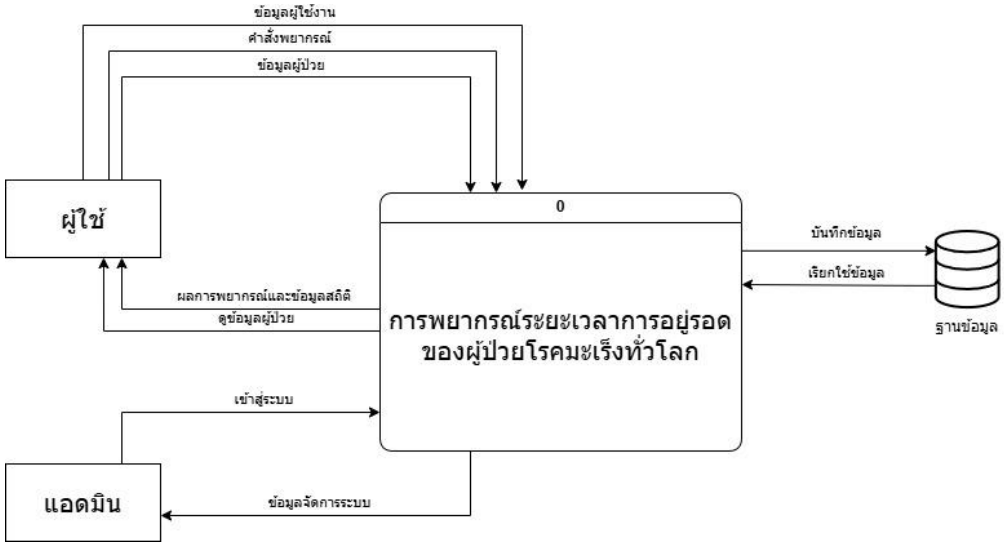
3.1.6 การนำตัวแบบไปใช้งานจริง (Deployment)

การนำโมเดลที่ผ่านการพัฒนา และเปรียบเทียบมาใช้งานจริง มีการดำเนินการดังต่อไปนี้

1) การเลือกตัวแบบที่เหมาะสม จากการวิเคราะห์ตัวแบบด้วยโปรแกรม RapidMiner ได้ทำการเปรียบเทียบค่าความ คลาดเคลื่อน (Error) ของตัวแบบทั้ง 3 ตัวแบบ และเลือกตัวแบบที่มีค่าความคลาดเคลื่อนต่ำ ที่สุดโดยใช้เกณฑ์ Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)

2) การพยากรณ์และประยุกต์ใช้งาน ตัวแบบที่ได้รับการเลือกจะถูกนำมาใช้ในการพยากรณ์ข้อมูลในอนาคต โดยอาศัย ข้อมูลจากชุดข้อมูลใหม่ เพื่อสร้างผลลัพธ์ที่มีความแม่นยำและสอดคล้องกับเป้าหมาย ที่กำหนดไว้

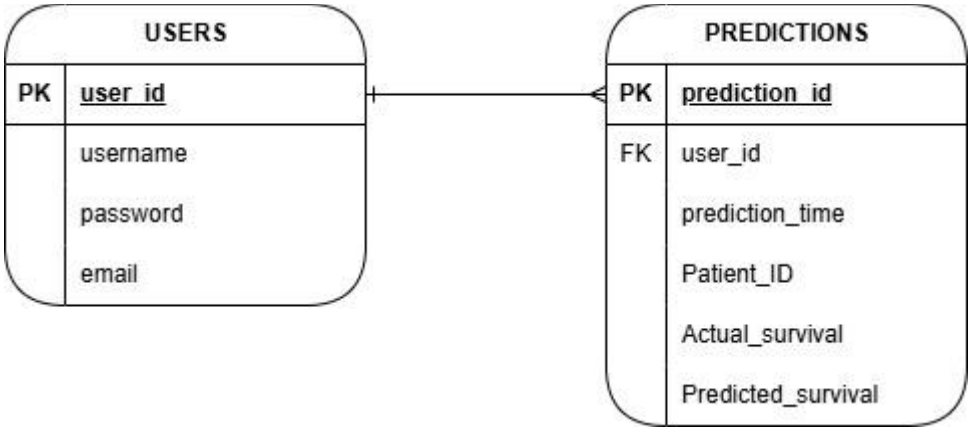
3.2 Context Diagram



ภาพที่ 3.3 ภาพของ Context Diagram

จากแผนภาพกระแสข้อมูลระดับ 0 (Context Diagram) แสดงภาพรวมการทำงานของระบบการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็ง มีส่วนเกี่ยวข้องกับระบบ 2 กลุ่มหลัก ได้แก่ ผู้ใช้งานทั่วไป (User) และผู้ดูแลระบบ (Admin) โดย ผู้ใช้งานทั่วไปจะทำการส่งข้อมูลการสมัครสมาชิก ข้อมูลการล็อกอินเข้าสู่ระบบ รวมถึงข้อมูลปัจจัยเสี่ยงและประวัติของผู้ป่วยเข้าสู่ระบบ เพื่อใช้เป็นคำสั่งพยากรณ์ จากนั้นระบบจะประมวลผลผ่านแบบจำลองโมเดลพยากรณ์ และส่งผลลัพธ์การพยากรณ์เป็นจำนวนปีที่รอดชีวิต พร้อมด้วยข้อมูลสถิติในรูปแบบกราฟ ในขณะเดียวกัน ผู้ดูแลระบบจะส่งข้อมูลล็อกอินเข้าสู่ระบบและมีสิทธิ์ส่งข้อมูลไฟล์ชุดข้อมูลผู้ป่วยแบบกลุ่ม (ไฟล์ CSV) เข้าสู่ระบบเพื่อปรับปรุงข้อมูล ซึ่งระบบจะทำการประมวลผลและส่งสถานะการนำเข้าข้อมูล หน้าการจัดการผู้ป่วย และหน้ากราฟสรุปสถิติภาพรวมกลับไปแสดงผลให้ผู้ดูแลระบบทราบ

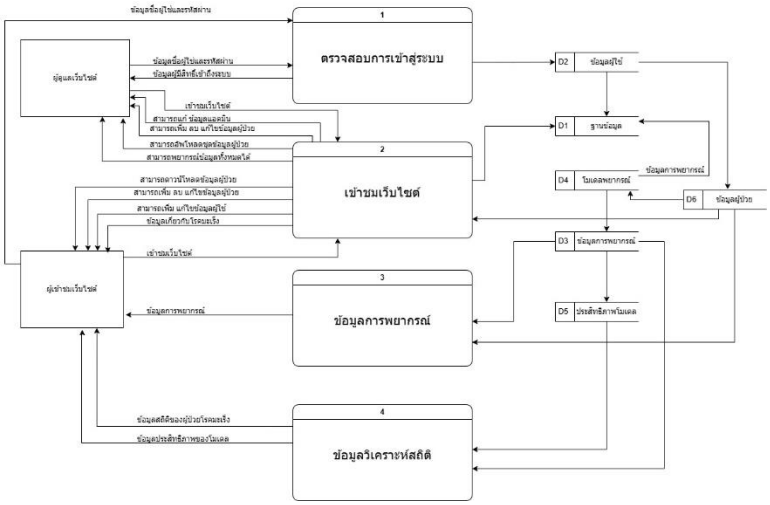
3.3 ER-Diagram



ภาพที่ 3.4 ภาพของ ER-Diagram

จากแผนภาพ โครงสร้างฐานข้อมูลของระบบประกอบด้วย 2 ตารางหลัก ได้แก่ ตาราง users สำหรับจัดเก็บข้อมูลบัญชีผู้ใช้ทั่วไปและบัญชีผู้ดูแลระบบ และตาราง Predictions ใช้สำหรับเก็บข้อมูลผลลัพธ์การพยากรณ์ เวลาพยากรณ์ ของผู้ป่วยแต่ละคน โดยแต่ละตารางจะมีความสัมพันธ์เชื่อมโยงกันผ่านรหัส USER (user_id) เพื่อให้ใช้สำหรับดึงข้อมูลการพยากรณ์

3.4 Data Flow Diagram Level 0



ภาพที่ 3.5 ภาพของ Data Flow Diagram Level 0

ตารางที่ 3.2 แสดงคำอธิบายการประมวลผลของโปรเซส 1.0 : ตรวจสอบการเข้าสู่ระบบ

Process Description	
System	การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการ เปรียบเทียบแบบจำลองที่เหมาะสมที่สุดเพื่อเผยแพร่ข้อมูลบนเว็บไซต์
DFD number	1.0
Process name	ตรวจสอบการเข้าสู่ระบบ
Input data flows	ข้อมูลผู้ใช้และรหัสผ่าน
Output data flows	ข้อมูลสิทธิ์การเข้าใช้งานระบบ
Data stores user	ข้อมูลผู้ใช้
Description	ตรวจสอบข้อมูลชื่อผู้ใช้และรหัสผ่านของผู้ใช้งานหรือผู้ดูแลระบบ โดยจะทำการ หากข้อมูลถูกต้องจะอนุญาตให้เข้าสู่ระบบและกำหนดสิทธิ์การใช้งานตามประเภทของผู้ใช้

ตารางที่ 3.3 แสดงคำอธิบายการประมวลผลของโปรเซส 2.0 : เข้าชมเว็บไซต์

Process Description	
System	การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการ เปรียบเทียบแบบจำลองที่เหมาะสมที่สุดเพื่อเผยแพร่ข้อมูลบนเว็บไซต์
DFD number	2.0
Process name	เข้าชมเว็บไซต์
Input data flows	สามารถเข้าชมเว็บไซต์
Output data flows	สามารถแก้ไข เพิ่มลบข้อมูลผู้ป่วย สามารถแก้ไขข้อมูลผู้ใช้ สามารถดาวน์โหลดข้อมูลผู้ป่วย ข้อมูลเกี่ยวกับโรคมะเร็ง สามารถพยากรณ์ข้อมูลทั้งหมด สามารถอัพโหลดชุดข้อมูลผู้ป่วย แก้ไขข้อมูลแอดมิน
Data stores user	ฐานข้อมูล ข้อมูลผู้ป่วย
Description	กระบวนการที่ผู้ใช้งานหรือผู้ดูแลระบบสามารถเข้าถึงระบบผ่านเว็บไซต์ เพื่อทำการจัดการข้อมูลต่าง ๆ เช่น การเพิ่ม แก้ไขหรือลบข้อมูลผู้ป่วย การอัปโหลดข้อมูลผู้ป่วยรวมถึงการส่งคำสั่งพยากรณ์ข้อมูลผู้ป่วย โดยระบบจะทำการเรียกใช้ข้อมูลจากฐานข้อมูลเพื่อนำมาแสดงผลให้ผู้ใช้งานสามารถเข้าถึงข้อมูลได้

ตารางที่ 3.4 แสดงคำอธิบายการประมวลผลของโปรเซส 3.0 : ข้อมูลการพยากรณ์

Process Description	
System	การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการ เปรียบเทียบแบบจำลองที่เหมาะสมที่สุดเพื่อเผยแพร่ข้อมูลบนเว็บไซต์
DFD number	3.0
Process name	ข้อมูลการพยากรณ์
Input data flows	ข้อมูลผู้ป่วย ข้อมูลโมเดลพยากรณ์
Output data flows	ข้อมูลการพยากรณ์
Data stores user	ข้อมูลการพยากรณ์ โมเดลพยากรณ์ ข้อมูลผู้ป่วย
Description	กระบวนการประมวลผลการพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็ง โดยระบบจะนำข้อมูลผู้ป่วยเข้าสู่แบบจำลองการพยากรณ์ที่จัดเก็บไว้ในระบบ เพื่อทำการคำนวณและสร้างผลลัพธ์การพยากรณ์ จากนั้นจะจัดเก็บข้อมูลผลการพยากรณ์และแสดงผลลัพธ์ให้ผู้ใช้งานสามารถตรวจสอบได้

ตารางที่ 3.5 แสดงคำอธิบายการประมวลผลของโปรเซส 4.0 : ข้อมูลวิเคราะห์สถิติ

Process Description	
System	การพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็งทั่วโลกด้วยการ เปรียบเทียบแบบจำลองที่เหมาะสมที่สุดเพื่อเผยแพร่ข้อมูลบนเว็บไซต์
DFD number	4.0
Process name	ข้อมูลวิเคราะห์สถิติ
Input data flows	ข้อมูลการพยากรณ์ ข้อมูลประสิทธิภาพของโมเดล
Output data flows	ข้อมูลสถิติของผู้ป่วยโรคมะเร็ง ข้อมูลประสิทธิภาพของโมเดล ข้อมูลการพยากรณ์
Data stores user	ข้อมูลการพยากรณ์ ประสิทธิภาพโมเดล
Description	กระบวนการวิเคราะห์และแสดงผลข้อมูลเชิงสถิติที่เกี่ยวข้องกับการพยากรณ์ เช่น จำนวนผู้ป่วย และค่าประสิทธิภาพของแบบจำลอง นำเสนอในรูปแบบกราฟหรือสถิติ เพื่อให้ผู้ใช้งานสามารถวิเคราะห์ข้อมูลได้ง่ายขึ้น

3.6 การออกแบบเว็บไซต์

3.7.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าแรกของเว็บไซต์แสดงเมนูต่างๆของหน้าเว็บ

ภาพที่ 3.6 แสดงโครงร่างการออกแบบหน้าแรกของเว็บไซต์

2) หน้าแสดงข้อมูล วัตถุประสงค์ ที่มาและความสำคัญ

ภาพที่ 3.7 แสดงโครงร่างการออกแบบหน้าที่มาและความสำคัญ

3) หน้า Dashboard แสดงผลการพยากรณ์ระยะเวลาอยู่รอดของผู้ป่วย

ภาพที่ 3.8 แสดงโครงร่างการออกแบบหน้าผลการวิเคราะห์ข้อมูล

4) หน้าแสดงข้อมูลผู้จัดทำ

ภาพที่ 3.9 แสดงโครงร่างการออกแบบหน้าข้อมูลผู้จัดทำ

5) หน้าเข้าสู่ระบบ

ภาพที่ 3.10 แสดงโครงร่างการออกแบบหน้าเข้าสู่ระบบ

3.6 บทสรุป

จากกระบวนการดำเนินงานในบทนี้ ได้อธิบายขั้นตอนการพัฒนาระบบพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วยโรคมะเร็ง โดยเริ่มจากการศึกษาและเตรียมข้อมูลผู้ป่วยให้เหมาะสมสำหรับการวิเคราะห์ จากนั้นนำกระบวนการ CRISP-DM มาใช้เป็นแนวทางในการดำเนินงาน ตั้งแต่การทำความเข้าใจข้อมูล การเตรียมข้อมูล การสร้างแบบจำลอง และการประเมินผลในการสร้างแบบจำลอง ผู้วิจัยได้ประยุกต์ใช้เทคนิค Machine Learning ได้แก่ Random Forest, Gradient Boosted Tree และ Neural Network เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง โดยใช้ตัวชี้วัด RMSE และ MAE ในการประเมินผล จากผลการทดลองพบว่า Random Forest ให้ค่าความคลาดเคลื่อนต่ำที่สุด จึงถูกเลือกเป็นแบบจำลองหลัก และนำไปประยุกต์ใช้ในระบบเว็บไซต์เพื่อพยากรณ์ระยะเวลาการอยู่รอดของผู้ป่วย พร้อมแสดงผลข้อมูลในรูปแบบ Dashboard เพื่อช่วยสนับสนุนการวิเคราะห์ได้อย่างมีประสิทธิภาพ