

บทที่ 3

วิธีดำเนินการวิจัย

3.1 ปัจจัยที่มีผลต่อการระบาดของโรคไข้เลือดออก

3.1.1 ศึกษางานวิจัยทั้งในและต่างประเทศเกี่ยวข้องกับการพยากรณ์โรคไข้เลือดออก และสรุปว่านักวิจัยใช้แอทริบิวต์ (ข้อมูล) อะไรบ้างเพื่อใช้พยากรณ์การระบาดของโรคไข้เลือดออก และทำการเปรียบเทียบ (ในรูปแบบตาราง) กับแอทริบิวต์ที่เราจะใช้ว่ามีอะไรที่เหมือนกัน และอะไรที่ต่างกัน

จากการศึกษาผลงานวิจัยของนักวิจัยหลายๆ ท่านทั้งในประเทศไทยและต่างประเทศ ผู้วิจัยสามารถจำแนกปัจจัยที่ใช้ในการพยากรณ์การระบาดของโรคไข้เลือดออก ดังนี้

1) ปัจจัยด้านลักษณะทางประชากร คือ ภูมิคุ้มกัน, อาชีพ, อุณหภูมิในร่างกาย, ปริมาณน้ำในร่างกาย และโรคประจำตัว

2) ปัจจัยด้านลักษณะของสภาพแวดล้อม คือ อุณหภูมิ, ปริมาณน้ำฝน, ความชื้นสัมพัทธ์, จำนวนผู้ป่วยโรคไข้เลือดออก, จำนวนผู้ป่วยไข้เดงกี, ลักษณะภูมิศาสตร์, อัตราแรงลม, ความสูงจากระดับน้ำทะเล, และฤดูกาล

ผู้วิจัยได้เปรียบเทียบปัจจัยที่ใช้ในการพยากรณ์การระบาดของโรคไข้เลือดออกระหว่างงานวิจัยของนักวิจัยหลายๆ ท่านกับข้อมูลที่ผู้วิจัยใช้ในการพยากรณ์ ดังตารางที่ 3-1

ตารางที่ 3-1 เปรียบเทียบปัจจัยที่ใช้ในการพยากรณ์โรคไข้เลือดออกระหว่างงานวิจัยอื่นๆ และงานวิจัยที่นำเสนอ

ปัจจัยที่ใช้ในการพยากรณ์โรคไข้เลือดออกจาก การทบทวนวรรณกรรม	ปัจจัยที่ใช้สำหรับงานวิจัยที่นำเสนอ
1.อุณหภูมิ	1.อุณหภูมิ
2.จำนวนประชากร	2.จำนวนประชากร
3.เพศ	3.ปริมาณน้ำฝน
4.อายุ	4.ความชื้นสัมพัทธ์
5.อาชีพ	5.จำนวนผู้ป่วยโรคไข้เลือดออก
6.อุณหภูมิในร่างกาย	6.แรงลม
7.ปริมาณน้ำในร่างกาย	7.ฤดูกาล

ตารางที่ 3-1 (ต่อ)

ปัจจัยที่ใช้ในการพยากรณ์โรคไข้เลือดออกจาก การทบทวนวรรณกรรม	ปัจจัยที่ใช้สำหรับงานวิจัยที่นำเสนอ
8.โรคประจำตัว	8.อัตราการติดเชื้อ Dengue ของลูกน้ำยุงลาย แยกตามฤดู
9.ปริมาณน้ำฝน	9.จำนวนยุงลายตัวเมียแยกตามฤดูกาล
10.ความชื้นสัมพัทธ์	10. อัตราการติดเชื้อ Dengue ชนิดต่างๆ ใน ผู้ป่วย
11.จำนวนผู้ป่วยโรคไข้เลือดออก	
12.จำนวนผู้ป่วยไข้เดงกี	
13.ลักษณะภูมิศาสตร์	
14.อัตราแรงลม	
15.ความสูงจากระดับน้ำทะเล	
16.ฤดูกาล	

จากการทบทวนวรรณกรรมผลงานวิจัยของนักวิจัยหลายๆ ท่านที่ได้ทำการศึกษาค้นคว้าพบว่ามี การเปรียบเทียบปัจจัยที่ใช้ในการพยากรณ์โดยใช้จำนวนผู้ป่วยเป็นตัวแปรตาม (Dependent variable) ส่วนตัวแปรต้น (Independent variables) ที่ใช้ในการพยากรณ์การระบาดของโรค ไข้เลือดออกนั้น นักวิจัยส่วนใหญ่ใช้ปัจจัยเพียงไม่กี่ปัจจัยในการพยากรณ์ อาทิเช่น อุณหภูมิกับ ปริมาณฝน หรือจำนวนประชากรกับฤดูกาล ดังนั้นในงานวิจัยนี้มีสมมติฐานสำคัญคือการใช้ปัจจัยใหม่ ได้แก่ อัตราการติดเชื้อของลูกน้ำยุงลายแยกตามฤดูและอัตราผู้ป่วยที่สงสัยว่าจะติดเชื้อจากผล ตรวจน้ำเหลืองของผู้ป่วย เพื่อให้การพยากรณ์การระบาดของโรคไข้เลือดออกมีความถูกต้องแม่นยำ และมีความน่าเชื่อถือในการพยากรณ์ ทั้งนี้จากการตรวจสอบเบื้องต้นพบว่า จำนวนยุงลายที่ติดเชื้อ แดงกี้ 1 และ 3 มีปริมาณสูง จะมีจำนวนผู้ป่วยโรคไข้เลือดออกสูงตามไปด้วย นอกจากนี้ยังสอดคล้อง กับผลงานวิจัยของ ชมพูศรี (Chompoosri, 2012) ที่พบว่าจำนวนผู้ป่วยไข้เลือดที่ติดเชื้อ Dengue 1 และ 3 มีมากกว่าซีโรทัยปี 2 และ 4

3.1.2 ทำการศึกษาวิธีการเก็บข้อมูลจากแหล่งข้อมูลต่างๆ ทำการวิเคราะห์ถึงความน่าเชื่อถือของวิธีการเก็บข้อมูลต่างๆ

การศึกษาถึงวิธีการเก็บข้อมูลจากแหล่งข้อมูลต่างๆ ที่จะนำมาใช้ในการพยากรณ์ ตารางที่ 3-2 ข้อมูล แหล่งข้อมูล และเครื่องมือที่ใช้ในการเก็บข้อมูล

ข้อมูล	แหล่งข้อมูล	เครื่องมือหรือวิธีเก็บข้อมูล
1. อุณหภูมิ	กรมอุตุนิยมวิทยา	ใช้เครื่องมือไฮโครมิเตอร์ (Dry-Wet Bulbs psychrometer) ในการวัดมีหน่วยเป็นองศาเซลเซียส
2. จำนวนประชากร	กรมการปกครอง	เก็บข้อมูลจำนวนประชากรรายตำบล รายอำเภอ และรายจังหวัด
3. ปริมาณน้ำฝน	กรมอุตุนิยมวิทยา	ใช้เครื่องวัดฝนแบบแก้วดวง ในการวัด มีหน่วยเป็นมิลลิเมตรต่อวัน
4. ความชื้นสัมพัทธ์	กรมอุตุนิยมวิทยา	ใช้เครื่องเทอร์โมไฮโกรกราฟ (Thermo-Hygrograph) ในการวัดมีหน่วยเป็นร้อยละ (%)
5. จำนวนผู้ป่วยโรคไข้เลือดออก	สำนักงานงานระบาดวิทยา	ระบบ รง. 506 สธ.
6. อัตราแรงลม	กรมอุตุนิยมวิทยา	ใช้เครื่องวัดความเร็วลม (Wind Speed Measurement) ในการวัดมีหน่วยเป็นกิโลเมตรต่อชั่วโมง
7. อัตราการติดเชื้อของลูกน้ำยุงลายแยกตามฤดู	ผศ.ดร.เผด็จ สิริยะเสถียร	นำยุงลายตัวเต็มวัยและลูกน้ำมาตรวจหาเชื้อด้วยวิธี RT-PCR โดยวิธี semi-nested PCR

3.1.3 ผลการวิเคราะห์เบื้องต้นจากข้อมูลที่รวบรวมได้

1) จำนวนยุงลายที่ติดเชื้อมากที่สุดในฤดูร้อน ซึ่งมีความสัมพันธ์กับจำนวนของผู้ป่วยในฤดูฝน หมายความว่าหากปีใดพบจำนวนยุงลายที่ติดเชื้อมาก ปีนั้นจะมีจำนวนผู้ป่วยโรคไข้เลือดออกแปรผันตรงกับจำนวนยุงลายที่ติดเชื้อมากเช่นกัน ดังนั้นผู้วิจัยจึงมีสมมติฐานว่า จำนวนยุงลายที่ติดเชื้อสามารถใช้เป็นปัจจัยที่สำคัญในการพยากรณ์การระบาดของโรคได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม

ตามงานวิจัยของ Chompoosri (2012) ได้ข้อสรุปว่าในฤดูร้อนนี้มีอัตราการกัดของยุงลายต่อหน่วยเวลา (นาทื) สูงกว่าฤดูอื่นๆ

2) จำนวนผู้ป่วยไข้เลือดออกสูงสุดอยู่ในฤดูฝน สันนิษฐานว่าเนื่องจากยุงลายที่ติดเชื้อใช้ระยะเวลาในการฟักตัวและเจริญเติบโตและออกหากินในช่วงฤดูฝนพอดี จึงทำให้เกิดการระบาดของไข้เลือดออกในช่วงฤดูฝนสูงกว่าฤดูร้อน และยังพบอีกว่าจำนวนยุงลายตัวเมียในฤดูฝนมีมากกว่าฤดูอื่นๆ ซึ่งสอดคล้องกับข้อเท็จจริงที่ว่ายุงตัวเมียนั้นที่กัดคน และเป็นสาเหตุสำคัญที่ทำให้เกิดการระบาดของโรค

3) ชนิดของเชื้อ Dengue ที่พบมากที่สุดคือ Dengue 1 และ 3 ผู้วิจัยตั้งสมมติฐานว่าอาจจะใช้เป็นข้อมูลสำคัญเพื่อพยากรณ์การระบาดของโรคไข้เลือดออกได้ นอกจากอุณหภูมิ ปริมาณน้ำฝน หรือจำนวนผู้ป่วย นอกจากนี้ยังสอดคล้องกับผลงานวิจัยของ จักรวาล ชมภูศรี (Jakkrawarn Chompoosri et al. 2012) ที่ว่าจำนวนผู้ป่วยไข้เลือดที่ติดเชื้อ Dengue 1 และ 3 มีมากกว่าซีโรทัยป์ 2 และ 4

3.2 การทำเหมืองข้อมูลและต้นแบบการพยากรณ์

ผู้วิจัยได้นำทฤษฎี CRISP-DM (บทที่ 2 รูป 2-1) มาใช้ในการดำเนินการจัดทำการใช้วิธีทางระบบธุรกิจอัจฉริยะสำหรับการป้องกันและควบคุมโรคไข้เลือดออก มีขั้นตอนดังนี้

1) ทำความเข้าใจปัญหา (Business Understanding)

ขั้นตอนการทำความเข้าใจวัตถุประสงค์ของงานวิจัย และทำความเข้าใจเกี่ยวกับการระบาดของไข้เลือดออก ระบุปัญหา ระบุปัจจัย และสถานะแวดล้อมที่ทำให้เกิดการระบาดของไข้เลือดออก มีกระบวนการย่อยดังนี้

- ความเข้าใจความต้องการของระบบโดยผู้วิจัยได้จัดทำแบบสอบถามผู้ใช้ระบบดังนี้

1. ต้องการให้ระบบพยากรณ์อะไรเกี่ยวกับโรคไข้เลือดออก

1) พยากรณ์จำนวนผู้ป่วยโรคไข้เลือดออก

2) พยากรณ์งบประมาณที่ควรจัดสรรในแต่ละพื้นที่

3) อื่นๆ

2. ต้องการให้ระบบรายงานอะไรเกี่ยวกับโรคไข้เลือดออก

- 1) รายงานพยากรณ์การระบาดของโรคไข้เลือดออกในปีถัดไป โดยแยกแต่ละฤดูกาล (ฤดูร้อน, ฤดูฝน และฤดูหนาว) แยกจังหวัด
- 2) รายงานพยากรณ์งบประมาณที่ใช้ในการรักษาโรคไข้เลือดออกเมื่อเกิดการระบาดของโรค
- 3) รายงานพยากรณ์จำนวนเวชภัณฑ์ที่ใช้ในการรักษาโรคไข้เลือดออกเมื่อเกิดการระบาดของโรค
- 4) รายงานพยากรณ์จำนวนบุคลากรทางการแพทย์ที่ใช้ในการรักษาโรคไข้เลือดออกเมื่อเกิดการระบาดของโรค
- 5) รายงานจำนวนผู้ป่วยโรคไข้เลือดออก แยกตามจังหวัด แยกตามฤดูกาล
- 6) รายงานจำนวนผู้ป่วยโรคไข้เลือดออก โดยแยกแต่ละฤดูกาล (ฤดูร้อน, ฤดูฝน และฤดูหนาว) แยกจังหวัดเปรียบเทียบจำนวนประชากรต่อแสนคนในแต่ละจังหวัด
- 7) รายงานการพยากรณ์พื้นที่ที่เสี่ยงต่อการระบาดของโรคไข้เลือดออก (สีเขียวปกติ, สีแดงเสี่ยง) โดยเปรียบเทียบจำนวนผู้ป่วยต่อจำนวนประชากรต่อแสนคนในแต่ละจังหวัด
 - ประเมินสถานการณ์ไข้เลือดออก
 - กำหนดเป้าหมายของการทำเหมืองข้อมูลเกี่ยวกับโรคไข้เลือดออก
 - วางแผนการดำเนินการเกี่ยวกับการจัดทำคลังข้อมูลโรคไข้เลือดออก

2) ทำความเข้าใจข้อมูล (Data Understanding)

จากการสอบถามข้อมูลเบื้องต้น และจากการวิจัยของที่ปรึกษาโครงการคือ รศ.นพ. เผด็จ สิริยะเสถียร พบว่าจังหวัดที่มีข้อมูลครบถ้วนได้แก่ นครปฐม ราชบุรี และสมุทรสาคร ดังนั้นขอบเขตการวิจัยนี้จะใช้ข้อมูลของ 3 จังหวัดดังกล่าวเป็นต้นแบบในการสร้างโมเดลการพยากรณ์โรคไข้เลือดออกก่อน อย่างไรก็ตามระบบจะถูกออกแบบไว้เพื่อรองรับการเพิ่มข้อมูลในอนาคต (Scalability) หากจะขยายระบบเพื่อพยากรณ์การระบาดในจังหวัดอื่นๆ ก็สามารถทำได้โดยการเพิ่มข้อมูลของปัจจัยต่างๆ ที่ใช้ในการพยากรณ์ของจังหวัดนั้นๆ ให้ครบถ้วน ระบบก็จะสามารถสร้างโมเดลการพยากรณ์ได้ใหม่อย่างอัตโนมัติโดยไม่ต้องทำการลื้อและสร้างระบบใหม่แต่อย่างใด ซึ่งมีขอบเขตดังนี้

(1) ขอบเขตด้านข้อมูล

แหล่งข้อมูลที่จะนำมาใช้ในการเป็นข้อมูลเข้า (input) ของระบบ สามารถรวบรวมจากแหล่งข้อมูลที่น่าเชื่อถือในประเทศไทย ได้แก่

1.1) กรมอุตุนิยมวิทยา ให้ข้อมูลด้านภูมิอากาศ อุณหภูมิ ความชื้นสัมพัทธ์ ปริมาณน้ำฝน

1.2) กองระบาดวิทยา กระทรวงสาธารณสุข ให้ข้อมูลเกี่ยวกับ จำนวนผู้ป่วยในแต่ละเดือน ย้อนหลังเป็นเวลาอย่างน้อย 5 ปี

1.3) งานวิจัยที่เกี่ยวข้องของ รศ.นพ. เผด็จ สิริยะเสถียร เช่น จำนวนปริมาณยุงลายที่ติดเชื้อในพื้นที่ 3 จังหวัด และบางอำเภอในแต่ละจังหวัด สมมติฐานการวิจัยสำคัญที่ตั้งไว้คือเป็นปัจจัยที่สำคัญอันหนึ่งที่มีผลต่ออัตราการแพร่ระบาดของโรคไข้เลือดออกและไม่เคยถูกนำมาใช้ในการสร้างโมเดลการพยากรณ์โรคไข้เลือดออกมาก่อน

แหล่งข้อมูลเหล่านี้มีการเก็บข้อมูลอย่างเป็นระบบโดยทำการเก็บตัวอย่างยุงในแต่ละฤดู และมีการตรวจหาเชื้อไข้เลือดออกโดยวิธี RT-PCR และมีความน่าเชื่อถือสูงที่สุดในปัจจุบันและอยู่ในประเทศไทย นักวิจัยสามารถติดต่อขอข้อมูลได้

(2) ขอบเขตด้านพื้นที่

กรณีศึกษาจังหวัดราชบุรี, นครปฐม และสมุทรสาคร

(3) ขอบเขตด้านการทำงานของระบบ

3.1) ระบบสามารถพยากรณ์ความเสี่ยงในการเกิดโรคระบาดในพื้นที่สำรวจได้อย่างน้อย 1 ปี (3 ฤดูกาล) ล่วงหน้า

3.2) ระบบสามารถแสดงข้อมูลในภาพรวมและข้อมูลเชิงลึกให้กับผู้ใช้เพื่อประกอบการตัดสินใจในเรื่องต่างๆ ที่เกี่ยวข้องได้ โดยการแสดงผลนี้มีทั้งส่วนที่เป็นตารางเพื่อให้ข้อเท็จจริง (Fact) กับผู้ใช้และแสดงเป็นแนวโน้มต่างๆ โดยใช้กราฟ (Graph) รูปแบบต่างๆ ได้

3.3) ระบบสามารถให้คำแนะนำเกี่ยวกับการจัดสรรทรัพยากรต่างๆ ในพื้นที่ตัวอย่างได้อย่างมีประสิทธิภาพ

3) เตรียมข้อมูล (Data preparation)

การ Data cleansing คือ การจัดการข้อมูลที่ผิดพลาด เช่น ข้อมูลบางส่วนหายไป หากไม่มีการทำขั้นตอนนี้จะทำให้โมเดลการพยากรณ์มีประสิทธิภาพที่ไม่ดีเท่าที่ควร ขั้นตอนนี้ถือเป็นขั้นตอน

ที่สำคัญและใช้เวลามากที่สุดของการทำโครงการวิจัยนี้ หลักการทำ Data cleansing ข้อมูลในโครงการวิจัยนี้มีหลักการคือ หากเป็นตัวเลข ระบบจะเติมข้อมูลที่หายไปโดยใช้ค่าเฉลี่ย (Average) ของข้อมูลในแอทริบิวต์นั้นๆ และหากเป็นตัวอักษร ระบบจะดูจากข้อมูลที่มีความถี่สูงสุด และเติมข้อมูลเข้าไปในช่องว่าง ข้อมูลที่ต้องทำ Data cleansing ในงานวิจัยนี้ ได้แก่ ปริมาณน้ำฝน จำนวนผู้ป่วยบางพื้นที่ๆ ไม่ได้มีการเก็บข้อมูล อัตราการติดเชื้อของยุงลาย เนื่องจากข้อมูลไม่ได้มีการเก็บอย่างครบถ้วน จึงไม่มีการทำ data cleansing

เมื่อตรวจสอบความถูกต้องของข้อมูลเสร็จเรียบร้อยแล้ว ขั้นตอนถัดไปของการจัดเตรียมข้อมูลคือการแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมจะประมวลผลหรือสร้างต้นแบบการพยากรณ์ ดังนั้นข้อมูลที่เก็บรวบรวมได้ถูกนำมาจัดเก็บในรูปแบบคลังข้อมูล ซึ่งประกอบด้วยตารางต่างๆ ดังต่อไปนี้

- 3.1) ข้อมูลอำเภอ (DimDistrict)
- 3.2) ข้อมูลตำบล (DimTambon)
- 3.3) ข้อมูลลักษณะภูมิประเทศ (DimGeoGraphy)
- 3.4) ข้อมูลเวชภัณฑ์ (DimMedicalSupplies)
- 3.5) ข้อมูลจังหวัด (DimProvince)
- 3.6) ข้อมูลภูมิภาค (DimRegional)
- 3.7) ข้อมูลเวลา (DimTime)
- 3.8) ข้อมูลผู้ใช้ (DimUser)
- 3.9) ข้อมูลงบประมาณ (FactBudget)
- 3.10) ข้อมูลจำนวนผู้ป่วยที่เสียชีวิต (FactDead)
- 3.11) ข้อมูลอัตราการป่วยตาย (FactDeadRateKey)
- 3.12) ข้อมูลชนิดเชื้อของโรคไข้เลือดออก (FactDengueType)
- 3.13) ข้อมูลความชื้นสัมพัทธ์ (FactHumidity)
- 3.14) ข้อมูลอัตราการติดเชื้อของยุงและลูกน้ำยุงลาย (FactLarvae)
- 3.15) ข้อมูลจำนวนบุคลากรทางการแพทย์ (FactMedicalPersonnel)
- 3.16) ข้อมูลจำนวนผู้ป่วยโรคไข้เลือดออก (FactPatient)
- 3.17) ข้อมูลอัตราการป่วยโรคไข้เลือดออก (FactPatientRate)

- 3.18) ข้อมูลจำนวนประชากร (FactPopulation)
- 3.19) ข้อมูลจำนวนสถานที่รักษา (FactPrimaryCareUnit)
- 3.20) ข้อมูลปริมาณฝน (FactRain)
- 3.21) ข้อมูลอุณหภูมิ (FactTemperature)
- 3.22) ข้อมูลอัตราแรงลม (FactWind)

4) สร้างแบบจำลอง (Modeling)

ขั้นตอนการสร้างแบบจำลองเพื่อนำไปวิเคราะห์ข้อมูลด้วยเทคนิคเหมือนข้อมูล ทั้งนี้ผู้วิจัยได้ใช้ทฤษฎี Microsoft Time Series Algorithm และ ทฤษฎีปัวซอง (Poisson regression analyses) ในการสร้างตัวทำนาย (prediction model) ทั้งนี้เนื่องจากทั้งสองวิธีการมีความเหมาะสมในการพยากรณ์ที่แตกต่างกัน สำหรับวิธี Microsoft Time Series Algorithm นั้นจะใช้เมื่อเราไม่ทราบค่าปัจจัยต่างๆ ที่จะมีผลต่อการระบาดของโรคไข้เลือดออกในอนาคต และสำหรับวิธีการปัวซองนั้นจะใช้เมื่อ ผู้ใช้ต้องการใส่ค่าประมาณการของปัจจัยต่างๆ ที่มีผลต่อการระบาดของโรคไข้เลือดออก เช่น อุณหภูมิ ความชื้นสัมพัทธ์ หรือร้อยละของยุงลายที่ติดเชื้อ เป็นต้น

5) ประเมิน (Evaluate)

การประเมินประสิทธิภาพของตัวแบบพยากรณ์โดยการตรวจสอบค่าความผิดพลาดของต้นแบบการพยากรณ์ (Error rate) จากข้อมูลจริงเพื่อเป็นตัวบ่งชี้ความน่าเชื่อถือของ Model ที่ได้ ในการศึกษาครั้งนี้ใช้เทคนิค Mean Absolute Percentage Error (MAPE) ในการประเมินประสิทธิภาพตัวแบบพยากรณ์

6) การนำไปใช้ (Deployment)

เป็นการนำผลลัพธ์หรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคข้อมูลไปใช้ประโยชน์ ตัวอย่างเช่น การรายงานพื้นที่เสี่ยงต่อการระบาดของไข้เลือดออกเพื่อเป็นการป้องกันและเฝ้าระวังการระบาดของโรคได้ทันทีและรวดเร็ว สามารถคำนวณงบประมาณ บุคลากรทางการแพทย์ และเวชภัณฑ์ที่ใช้ในการรักษาผู้ป่วยโรคไข้เลือดออก เป็นต้น