

บทที่ 5

สรุป และขอเสนอแนะ

จากผลการวิเคราะห์ข้อมูลเพื่อทำนายลักษณะของผู้ป่วยโรคเบาหวานจากชุดข้อมูลทุติยภูมิแบบเปิด และการจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูล ทางผู้วิเคราะห์ได้ทำการสรุปผลการทำโครงการ ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่างๆ ของการทำโครงการ และขอเสนอแนะเพื่อพัฒนาโครงการ ดังนี้

5.1 บทสรุปผลโครงการ

ผู้วิเคราะห์ข้อมูลเพื่อทำนายลักษณะของผู้ป่วยโรคเบาหวานจากชุดข้อมูลทุติยภูมิแบบเปิด จากเว็บไซต์ [Kaggle – 100000 Diabetes Clinical Dataset](#) และได้ทำการวิเคราะห์ข้อมูลด้วย การทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายลักษณะของผู้ป่วยโรคเบาหวาน และสามารถนำไปทำนายผลด้วยต้นไม้ตัดสินใจจากเทคนิคการจำแนกประเภทข้อมูล (Classification) เพื่อให้ได้ผลการทำนายที่เหมาะสม และผลการคำนวณค่าความถูกต้อง (accuracy) ค่าเรียกกลับ (recall) และค่าความแม่นยำ (precision) โดยข้อมูลแบ่งเป็นชุดฝึกและทดสอบ โดยกำหนดอัตราส่วนเป็น 70 : 30 ซึ่งหมายความว่า 70% ของข้อมูลจะถูกนำไปใช้สำหรับการฝึกโมเดล อีก 30% ใช้สำหรับการประเมินผลการทำงานของโมเดล เพื่อป้องกันการเกิด overfitting และช่วยให้สามารถประเมินประสิทธิภาพของโมเดลกับข้อมูลที่ไม่เคยเห็นมาก่อนได้อย่างเหมาะสม พร้อมทั้งเลือกวิธีการสุ่มตัวอย่างแบบ linear sampling เพื่อจัดเรียงข้อมูลตามลำดับที่กำหนด ซึ่งชุดฝึกผ่านการปรับสมดุลด้วย SMOTE Upsampling เพื่อปรับชุดข้อมูลที่มีหลายคลาสแบบไม่สมดุล (imbalanced dataset) ให้สมดุล และลดความลำเอียง (bias) และจากการวัดประสิทธิภาพของโมเดล (Evaluation) การทดสอบประสิทธิภาพของแบบจำลองได้ดำเนินการวัดค่าความแม่นยำ ค่าเรียกกลับ และค่าความถูกต้อง เพื่อสร้างแบบจำลองการพยากรณ์จากการเรียนรู้ และทดสอบแบบจำลองการ พยากรณ์โดยค่าความถูกต้อง (Accuracy) โดยผลการคำนวณค่าความถูกต้องที่ได้จากการ คำนวณของโปรแกรม Rapid Miner มีค่าความถูกต้องเฉลี่ยเท่ากับ 90.69% ซึ่งมีผลอยู่ในระดับดีมาก ทางผู้วิเคราะห์ข้อมูล

นำผลลัพธ์ที่ทำการวิเคราะห์แล้ว นำมาสร้างเป็นสารสนเทศจัดแสดงผ่านเว็บไซต์เพื่อเผยแพร่ข้อมูลการวิเคราะห์ลักษณะของผู้ป่วยโรคเบาหวานจากชุดข้อมูลหุติยภูมิแบบเปิด โดยใช้ชุดคำสั่ง HTML CSS และ PHP ร่วมกับการนำเสนอข้อมูลแบบ Visualization ที่ประกอบด้วยแผนภูมิแท่ง แผนภูมิวงกลม เป็นต้น ด้วยโปรแกรม Power BI และ เผยแพร่ผลการวิเคราะห์ลักษณะของผู้ป่วยโรคเบาหวาน ด้วยการพัฒนาเว็บไซต์เพื่อเผยแพร่ให้ผู้สนใจได้รับทราบข้อมูลผลการวิเคราะห์

5.2 ข้อจำกัดของระบบ

5.2.1 ผู้ใช้งานทั่วไปไม่สามารถเปลี่ยนแปลงรูปแบบการแสดงผลข้อมูลแผนภาพ (visualization) ของเว็บไซต์ได้ เนื่องจากผู้วิเคราะห์ข้อมูลจัดทำแผนภาพในโปรแกรม Power BI โดยใช้ Account ของผู้วิเคราะห์ข้อมูลเอง จากนั้นได้นำไปฝังในเว็บไซต์ที่แสดงผล ดังนั้นผู้ใช้งานจึงไม่สามารถเปลี่ยนแปลงรูปแบบการแสดงผลข้อมูลอื่น ๆ ได้ เช่น หากผู้ใช้งานข้อมูลแบบกราฟแท่งแล้วอยากเปลี่ยนไปดูข้อมูลนี้ในรูปแบบกราฟอื่นๆ จะไม่สามารถเปลี่ยนแปลงได้ แต่ผู้ใช้งานสามารถกรองดูข้อมูลผ่านบนเว็บไซต์ได้

5.2.2 ข้อมูลที่ใช้ในเว็บไซต์เป็นข้อมูลหุติยภูมิจากชุดข้อมูลบนเว็บไซต์ Kaggle ซึ่งไม่ได้อัปเดตแบบเรียลไทม์ ทำให้ผลการวิเคราะห์และแสดงผลอาจไม่สะท้อนสถานการณ์ผู้ป่วยโรคเบาหวานในปัจจุบันได้อย่างครบถ้วน

5.2.3 ข้อจำกัดของการเข้าถึงข้อมูลผู้ใช้ไม่สามารถ เพิ่ม ลบ หรือแก้ไขข้อมูลภายในเว็บไซต์ได้ เนื่องจากถูกจำกัดการเข้าถึงได้เฉพาะแอดมินที่สามารถเข้าสู่ระบบเพื่อแก้ไข เพิ่ม ลบ หรือแก้ไขข้อมูลในส่วนนี้ แต่ผู้ใช้งานสามารถดาวน์โหลดชุดข้อมูลที่ใช้ในการวิเคราะห์ได้จากหน้าแสดงผล Dashboard

5.2.4 ผู้ใช้งานไม่สามารถรับรู้ถึงการแจ้งเตือนการปรับปรุงข้อมูลของเว็บไซต์ หากเว็บไซต์เกิดการปรับปรุงข้อมูลใหม่ ผู้ใช้งานจะไม่สามารถรับรู้ได้

5.3 ปัญหาและอุปสรรคของโครงการ

5.3.1 ผู้วิเคราะห์ข้อมูลไม่สามารถจัดเก็บรวบรวมข้อมูลด้วยตนเองได้ เนื่องจากชุดข้อมูลปัญหาโรคเบาหวาน ต้องอาศัยแพทย์ผู้เชี่ยวชาญเฉพาะทางโดยตรงในการให้คำปรึกษา

เกี่ยวกับข้อมูล จึงต้องพึ่งข้อมูลชุดเดียวจาก Kaggle ทำให้การวิเคราะห์เชิงลึกและการตีความทำได้จำกัด

5.3.2 ข้อจำกัดของข้อมูลที่ไม่ได้ทำการจัดเก็บด้วยตนเอง จึงมีข้อมูลที่ไม่ได้อัปเดตแบบเรียลไทม์และมีความครอบคลุมของข้อมูลที่ไม่เพียงพอ

5.3.3 การจัดวางแผนภาพให้อยู่ในรูปแบบที่เข้าใจได้ง่าย เพราะข้อมูลที่วิเคราะห์ออกมาหลายประเด็น

5.3.4 สัดส่วนผู้ป่วยเบาหวาน (diabetes=1) มีน้อยเมื่อเทียบกับกลุ่มปกติ ทำให้ตัวชี้วัดสำหรับแยกกิ่งของโมเดล มีอคติและอาจเลือกตัวแปรที่ไม่สะท้อนเหตุผลจริง อีกทั้งตัวแปรเชิงหมวดหมู่บางตัว เช่น smoking_history ถูกเข้ารหัสเป็นตัวเลข อาจถูกโมเดลตีความเป็นลำดับเชิงปริมาณ อาจส่งผลให้การเลือก Root Node และเกณฑ์ตัดสินใจผิดพลาดและเสี่ยงต่อการตีความผิด

5.3.5 เนื่องจากข้อมูลที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลที่ไม่สมบูรณ์ หรือจัดอยู่ในรูปแบบที่ไม่สามารถนำไปประมวลผลได้ทันที ต้องทำการจัดรูปแบบของข้อมูล หรือทำความสะอาดข้อมูล ซึ่งอาจทำให้ข้อมูลบางส่วนที่มีความซ้ำซ้อนมาก ตัวอย่างเช่น แอดทริบิวต์ smoking_history ที่มีความหมายข้อมูลคล้ายกันในคอลัมน์ของแต่ละประเภท จึงได้ทำการแปลงข้อความให้อยู่ในรูปแบบที่สามารถเข้าใจได้ตรงกัน และผู้วิเคราะห์ข้อมูลได้ลบข้อมูลบางส่วนที่ไม่สามารถนำมาวิเคราะห์ได้ ดังนั้นในการวิเคราะห์ข้อมูลในครั้งถัดไป ควรจัดเก็บรวบรวมชุดข้อมูลด้วยตนเองเพื่อให้ข้อมูลมีความถูกต้องสูง

5.4 ข้อเสนอแนะ

การวิเคราะห์ข้อมูลเพื่อทำนายลักษณะของผู้ป่วยโรคเบาหวานจากชุดข้อมูลชุดเดียวแบบเปิด สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ที่จัดทำเสร็จสิ้นนี้แม้จะสามารถแสดงผลข้อมูลได้ถูกต้องตรงตามวัตถุประสงค์ และขอบเขตที่ตั้งไว้แต่ก็ยังมีข้อจำกัดหลายประการ ซึ่งหากจะพัฒนาให้เว็บไซต์เผยแพร่ข้อมูลนี้แสดงผลข้อมูล และทำงานได้อย่างมีประสิทธิภาพมากขึ้น ผู้พัฒนาควรปรับปรุงในส่วนต่างๆ ดังต่อไปนี้

5.4.1 ควรมีการจัดเก็บรวบรวมข้อมูลผู้ป่วยโรคเบาหวานเพิ่มเติมจากแหล่งข้อมูลจริง เพื่อให้ได้ข้อมูลที่ครอบคลุมกลุ่มเป้าหมายและมีความทันสมัยมากขึ้น เช่น การบันทึกข้อมูล

จากสถานพยาบาลในระดับจังหวัดหรือโรงพยาบาล เพื่อใช้เปรียบเทียบกับข้อมูลทุติยภูมิจาก Kaggle จะช่วยให้ผลการวิเคราะห์แนวโน้มและปัจจัยเสี่ยงของโรคเบาหวานมีความถูกต้องและสอดคล้องกับสถานการณ์ปัจจุบันมากขึ้น

5.4.2 ในการวิเคราะห์ข้อมูลในครั้งต่อไป ควรทำการศึกษาข้อมูลในเชิงลึก และปรึกษาแพทย์ผู้เชี่ยวชาญเฉพาะทางโดยตรง เพื่อความถูกต้องในการวิเคราะห์ข้อมูล และช่วยในการเลือกตัวแปรที่มีความสำคัญต่อภาวะเบาหวาน

5.4.3 ควรพัฒนาเว็บไซต์ให้สามารถอัปเดตข้อมูลอัตโนมัติได้เมื่อมีการเพิ่มชุดข้อมูลใหม่ในฐาน Kaggle เพื่อให้แดชบอร์ดสามารถแสดงผลข้อมูลล่าสุดได้แบบเรียลไทม์ ช่วยลดข้อจำกัดของข้อมูลที่ไม่เป็นปัจจุบัน และเพิ่มประสิทธิภาพในการติดตามแนวโน้มของผู้ป่วยโรคเบาหวานอย่างต่อเนื่อง

5.4.4 ในการวิเคราะห์ข้อมูลในครั้งต่อไป เว็บไซต์ที่เผยแพร่ข้อมูลควรทำเมนูแจ้งเตือน เพื่อให้ผู้ใช้สามารถรู้ถึงการปรับปรุงข้อมูลครั้งล่าสุดของเว็บไซต์ หากเว็บไซต์เกิดการปรับปรุงข้อมูลใหม่ เพื่อให้เกิดความน่าเชื่อถือของข้อมูลหากผู้ใช้นำข้อมูลสารสนเทศนี้ไปใช้งานต่อไป

5.4.5 ในการวิเคราะห์ข้อมูลในครั้งต่อไป ควรปรึกษาแพทย์ หรือผู้เชี่ยวชาญเฉพาะทางโดยตรง เพื่อความถูกต้องในการวิเคราะห์ข้อมูล