

## บทที่ 5

### สรุป อภิปรายผล และข้อเสนอแนะ

#### 5.1 สรุป

การทำเหมืองข้อมูล (Data Mining) หรือการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases – KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) ของจากข้อมูลจำนวนมากโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบหรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูลที่มีจำนวนมากเพื่อค้นหารูปแบบแนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ ซึ่งเป็นเทคโนโลยีใหม่ของการประยุกต์ใช้ข้อมูลที่เก็บอยู่ในฐานข้อมูล เพื่อให้เกิดประโยชน์สูงสุดในการใช้ข้อมูล การประยุกต์ใช้ข้อมูลที่กล่าวถึงนี้มีได้หลายแนวทาง แต่โดยทั่วไปมักจะเป็นการสรุปภาพรวมของข้อมูลในฐานข้อมูล การวิเคราะห์แนวโน้ม พฤติกรรมการเปลี่ยนแปลงของข้อมูล หรือ การค้นหาความสัมพันธ์ที่ซ่อนอยู่ภายในกลุ่มของข้อมูล

งานวิจัยนี้มุ่งเน้นการค้นหาเทคนิคด้านเหมืองข้อมูล เพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ โดยมีจุดมุ่งหมายที่จะทดสอบประสิทธิภาพของอัลกอริทึมต่างๆ ที่ใช้ในการทำเหมืองข้อมูล มุ่งเน้นการจำแนกข้อมูล (Classification) เพื่อค้นหาอัลกอริทึมที่เหมาะสมที่สุดสำหรับฐานข้อมูลทางการแพทย์ โดยใช้อัลกอริทึมพื้นฐาน 7 อัลกอริทึม ซึ่งประกอบด้วย Naïve Bayes, Multilayer Perceptron, Radial Basis Function Network, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Ripper รวมถึงศึกษาเปรียบเทียบวิธีลดคุณลักษณะที่เหมาะสมด้วยวิธี Correlation-based Feature Subset Selection (CFS) และวิธี Feature selection method based on correlation measure and relevance & redundancy analysis (FCBF) รวมถึงทดสอบกับอัลกอริทึมประเภท Single learning กับ Multiple learning โดยเพิ่มประสิทธิภาพด้วยวิธี Bagging และ Boosting โดยทดสอบกับข้อมูลทางการแพทย์ทั้ง 13 ชุด ซึ่งประกอบด้วย breast-cancer, breast-w, diabetes, heart-c, heart-statlog, hepatitis, hypothyroid, leukemia, liver-disorders, lung-cancer, lymphography, postoperative-patient, primary-tumor ซึ่งผลการทดลองสามารถสรุปได้ดังนี้

### 5.1.1. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพแต่ละอัลกอริทึม โดยไม่ลดคุณลักษณะ

เมื่อเรียนรู้กลุ่มตัวอย่างโดยไม่ลดคุณลักษณะ ตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่นำเสนอ โดยวัดประสิทธิภาพจากค่าความความถูกต้อง (Accuracy) พบว่า กลุ่มข้อมูล breast-cancer อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 75.52% กลุ่มข้อมูล breast-w อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.99% กลุ่มข้อมูล diabetes อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 77.34% กลุ่มข้อมูล heart-c อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.15% กลุ่มข้อมูล heart-statlog อัลกอริทึม Radial Basis Function กับ Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.07% กลุ่มข้อมูล hepatitis อัลกอริทึม Radial Basis Function กับ K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.80% กลุ่มข้อมูล hypothyroid อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 99.57% กลุ่มข้อมูล leukemia อัลกอริทึม Naive Bayes กับ Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 98.61% กลุ่มข้อมูล liver-disorders อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 68.69% กลุ่มข้อมูล lung-cancer อัลกอริทึม Naive Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 62.5% กลุ่มข้อมูล lymphography อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 86.48% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor กับ Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 71.11% กลุ่มข้อมูล primary-tumor อัลกอริทึม Naive Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 50.14%

### 5.1.2. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพแต่ละอัลกอริทึมเมื่อเรียนรู้ด้วยวิธี Bagging

เมื่อเรียนรู้กลุ่มตัวอย่างโดยไม่ลดคุณลักษณะด้วยวิธี Bagging ตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่นำเสนอ โดยวัดประสิทธิภาพจากค่าความความถูกต้อง (Accuracy) พบว่า กลุ่มข้อมูล breast-cancer อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 74.82% กลุ่มข้อมูล breast-w อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.85% กลุ่มข้อมูล diabetes อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 77.47% กลุ่มข้อมูล heart-c อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.14% กลุ่มข้อมูล heart-statlog อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.44% กลุ่มข้อมูล hepatitis อัลกอริทึม Radial

Basis Function ให้ประสิทธิภาพออกมาดีที่สุดคือ 87.09% กลุ่มข้อมูล hypothyroid อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 99.57% กลุ่มข้อมูล leukemia อัลกอริทึม Naive Bayes, Radial Basis Function กับ Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 97.22% กลุ่มข้อมูล liver-disorders อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 72.75% กลุ่มข้อมูล lung-cancer อัลกอริทึม Naive Bayes, Multilayer Perceptron กับ K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 53.12% กลุ่มข้อมูล lymphography อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.81% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 71.11% กลุ่มข้อมูล primary-tumor อัลกอริทึม Naive Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 49.55%

### 5.1.3. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพแต่ละอัลกอริทึมเมื่อเรียนรู้ด้วยวิธี Boosting

เมื่อเรียนรู้กลุ่มตัวอย่างโดยไม่ลดคุณลักษณะด้วยวิธี Boosting ตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่น่าเสนอ โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) พบว่า กลุ่มข้อมูล breast-cancer อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 73.42% กลุ่มข้อมูล breast-w อัลกอริทึม Support Vector Machine กับ K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.70% กลุ่มข้อมูล diabetes อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 77.34% กลุ่มข้อมูล heart-c อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.81% กลุ่มข้อมูล heart-statlog อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.07% กลุ่มข้อมูล hepatitis อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 87.09% กลุ่มข้อมูล hypothyroid อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 99.62% กลุ่มข้อมูล leukemia อัลกอริทึม Naive Bayes กับ Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 98.61% กลุ่มข้อมูล liver-disorders อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 73.33% กลุ่มข้อมูล lung-cancer อัลกอริทึม Support Vector Machine กับ Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 50% กลุ่มข้อมูล lymphography อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.45% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 71.11% กลุ่มข้อมูล primary-tumor อัลกอริทึม Naive Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 50.14%

#### 5.1.4. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพเมื่อลดคุณลักษณะด้วยวิธี CFS

เมื่อทำการลดคุณลักษณะด้วย CFS และทำการเรียนรู้กลุ่มตัวอย่างตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่นำเสนอ โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) พบว่า กลุ่มข้อมูล breast-cancer อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 73.77% กลุ่มข้อมูล breast-w อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.99% กลุ่มข้อมูล diabetes อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 76.82% กลุ่มข้อมูล heart-c อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.48% กลุ่มข้อมูล heart-statlog อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.55% กลุ่มข้อมูล hepatitis อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 87.74% กลุ่มข้อมูล hypothyroid อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 97.16% กลุ่มข้อมูล leukemia อัลกอริทึม Radial Basis Function Network ให้ประสิทธิภาพออกมาดีที่สุดคือ 100% กลุ่มข้อมูล liver-disorders อัลกอริทึม Multilayer Perceptron กับ Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 62.02% กลุ่มข้อมูล lung-cancer อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 75% กลุ่มข้อมูล lymphography อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.81% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor กับ Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 71.11% กลุ่มข้อมูล primary-tumor อัลกอริทึม Naïve Bayes กับ K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 47.78%

#### 5.1.5. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพเมื่อลดคุณลักษณะด้วยวิธี FCBF

เมื่อทำการลดคุณลักษณะด้วย FCBF และทำการเรียนรู้กลุ่มตัวอย่างตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่นำเสนอ โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) พบว่ากลุ่มข้อมูล breast-cancer อัลกอริทึม Support Vector Machine กับ Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 72.37% กลุ่มข้อมูล breast-w อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.99% กลุ่มข้อมูล diabetes อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 77.47% กลุ่มข้อมูล heart-c อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.48% กลุ่มข้อมูล heart-statlog อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.44% กลุ่มข้อมูล hepatitis อัลกอริทึม Multilayer Perceptron ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.80% กลุ่มข้อมูล hypothyroid อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 97.53% กลุ่มข้อมูล leukemia อัลกอริทึม Multilayer Perceptron

กับ Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 100% กลุ่มข้อมูล liver-disorders อัลกอริทึม Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 62.02% กลุ่มข้อมูล lung-cancer อัลกอริทึม K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 65.62% กลุ่มข้อมูล lymphography อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 81.75% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor กับ Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 71.11% กลุ่มข้อมูล primary-tumor อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 47.19%

#### 5.1.6. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพวิธี Single learning กับ Multiple learning

เมื่อทำการเปรียบเทียบประสิทธิภาพด้าน Accuracy ของ Single learning กับ Multiple learning จากการเรียนรู้กลุ่มตัวอย่างตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่นำเสนอ โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) พบว่ากลุ่มข้อมูล breast-cancer อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 75.52% กลุ่มข้อมูล breast-w อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.99% กลุ่มข้อมูล diabetes อัลกอริทึม Bagging Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 77.47% กลุ่มข้อมูล heart-c อัลกอริทึม Bagging Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 85.14% กลุ่มข้อมูล heart-statlog อัลกอริทึม Bagging Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 84.44% กลุ่มข้อมูล hepatitis อัลกอริทึม Bagging RBF กับ Boosting J48 ให้ประสิทธิภาพออกมาดีที่สุดคือ 87.09% กลุ่มข้อมูล hypothyroid อัลกอริทึม Boosting Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 99.62% กลุ่มข้อมูล leukemia อัลกอริทึม Naïve Bayes ,Boosting Naïve Bayes กับ Support Vector Machine ,Boosting Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 98.61% กลุ่มข้อมูล liver-disorders อัลกอริทึม Boosting Ripper ให้ประสิทธิภาพออกมาดีที่สุดคือ 73.33% กลุ่มข้อมูล lung-cancer อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 62.5% กลุ่มข้อมูล lymphography อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 86.48% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor ,Ripper, Bagging และ Boosting K-Nearest Neighbor ให้ประสิทธิภาพออกมาดีที่สุดคือ 71.11% กลุ่มข้อมูล primary-tumor อัลกอริทึม Naïve Bayes กับ Boosting Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 50.14%

### 5.1.7. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการลดคุณลักษณะด้วยวิธี CFS และ FCBF

เมื่อทำการเปรียบเทียบประสิทธิภาพด้าน Accuracy เมื่อไม่ลดคุณลักษณะกับลดคุณลักษณะด้วยวิธี CFS และวิธี FCBF จากการเรียนรู้กลุ่มตัวอย่างตามแบบจำลองการวิเคราะห์โรคอัตโนมัติที่นำเสนอ โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) พบว่า

กลุ่มแรกการลดคุณลักษณะไม่ส่งผลให้ค่าความถูกต้องเพิ่มขึ้น มีดังนี้ กลุ่มข้อมูล breast-cancer อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 75.52% กลุ่มข้อมูล breast-w อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 96.99% กลุ่มข้อมูล hypothyroid อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 99.57% กลุ่มข้อมูล liver-disorders อัลกอริทึม Decision Tree ให้ประสิทธิภาพออกมาดีที่สุดคือ 68.69% กลุ่มข้อมูล lymphography อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพออกมาดีที่สุดคือ 86.48% และกลุ่มข้อมูล primary-tumor อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพออกมาดีที่สุดคือ 50.14%

กลุ่มสองการลดคุณลักษณะสามารถลดขนาดของมิติของข้อมูลได้เป็นอย่างดีแล้ว ยังส่งผลให้ค่าความถูกต้องเพิ่มมากขึ้นมีดังนี้ กลุ่มข้อมูล diabetes อัลกอริทึม CFS Naïve Bayes กับ FCBF Naïve Bayes ให้ประสิทธิภาพสูงขึ้นเป็น 77.47% กลุ่มข้อมูล heart-c อัลกอริทึม CFS Naïve Bayes กับ FCBF Naïve Bayes ให้ประสิทธิภาพออกมาสูงขึ้นเป็น 84.48% กลุ่มข้อมูล heart-statlog อัลกอริทึม CFS Naïve Bayes ให้ประสิทธิภาพสูงขึ้นเป็น 85.55% กลุ่มข้อมูล hepatitis อัลกอริทึม CFS Naïve Bayes ให้ประสิทธิภาพสูงขึ้นเป็น 87.74% กลุ่มข้อมูล leukemia อัลกอริทึม CFS RBF,FCBF Multilayer Perceptron และ FCBF Support Vector Machine ให้ประสิทธิภาพสูงขึ้นเป็น 100% กลุ่มข้อมูล lung-cancer อัลกอริทึม CFS Support Vector Machine ให้ประสิทธิภาพสูงขึ้นเป็น 75% กลุ่มข้อมูล postoperative-patient อัลกอริทึม K-Nearest Neighbor,Ripper และ CFS K-Nearest Neighbor,CFS Ripper เมื่อลดคุณลักษณะแล้วให้ประสิทธิภาพออกมาเท่าเดิมคือ 71.11% จากข้อมูลกราฟดังกล่าวทำให้เราสามารถสรุปได้ว่า อัลกอริทึมประเภทไหนเหมาะสมกับชุดข้อมูลแบบใดและการลดคุณลักษณะประเภทใดส่งผลให้ค่าความถูกต้องเพิ่มขึ้นสูงสุด

## 5.2 อภิปรายผล

ผลจากการทดลองการค้นหาเทคนิคด้านเหมืองข้อมูล เพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ โดยมีจุดมุ่งหมายที่จะทดสอบประสิทธิภาพของอัลกอริทึมที่ใช้ในการจำแนกข้อมูล (Classification) เพื่อค้นหาอัลกอริทึมที่เหมาะสมที่สุดสำหรับฐานข้อมูลทางการแพทย์ สามารถอภิปรายผลได้ □ ดังนี้

5.2.1 โมเดลการวิเคราะห์โรคอัตโนมัติที่สร้างได้ มีประสิทธิภาพความถูกต้องในการจำแนกประเภทของโรคในระดับ 80 % ขึ้นไป โดยไม่ลดคุณลักษณะ ทำการเรียงตามค่าความถูกต้อง (Accuracy) ประกอบด้วย กลุ่มข้อมูล hypothyroid สร้างโมเดลด้วยอัลกอริทึม Decision Tree ให้ประสิทธิภาพ 99.57% กลุ่มข้อมูล leukemia สร้างโมเดลด้วยอัลกอริทึม Naive Bayes กับ Support Vector Machine ให้ประสิทธิภาพ 98.61% กลุ่มข้อมูล breast-w สร้างโมเดลด้วยอัลกอริทึม Support Vector Machine ให้ประสิทธิภาพ 96.99% กลุ่มข้อมูล lymphography สร้างโมเดลด้วยอัลกอริทึม Support Vector Machine ให้ประสิทธิภาพ 86.48% กลุ่มข้อมูล hepatitis สร้างโมเดลด้วยอัลกอริทึม Radial Basis Function กับ K-Nearest Neighbor ให้ประสิทธิภาพ 85.80% กลุ่มข้อมูล heart-c สร้างโมเดลด้วยอัลกอริทึม Support Vector Machine ให้ประสิทธิภาพ 84.15% กลุ่มข้อมูล heart-statlog สร้างโมเดลด้วยอัลกอริทึม Radial Basis Function กับ Support Vector Machine ให้ประสิทธิภาพ 84.07% ตามลำดับ ซึ่งโมเดลดังกล่าวมีคุณภาพในระดับที่ยอมรับได้ตามเกณฑ์ 80 % สามารถโมเดลดังกล่าวไปพัฒนาเป็นซอฟต์แวร์ในการวินิจฉัยโรคอัตโนมัติได้

5.2.2 การเพิ่มประสิทธิภาพความถูกต้อง (Accuracy) ด้วยวิธี Multiple learning โดยใช้ อัลกอริทึม Bagging และ Boosting นั้น ส่งผลให้ค่าความถูกต้อง เพิ่มขึ้นเฉพาะบางกลุ่มข้อมูลเท่านั้น โดยมีข้อสังเกตว่า สัดส่วนของกลุ่มตัวอย่างในแต่ละคลาสจะต้องมีปริมาณใกล้เคียงกัน หรือเท่ากัน กรณีที่การกระจายของของกลุ่มตัวอย่างในแต่ละคลาส มีสัดส่วนที่แตกต่างกันมาก ส่งผลให้เทคนิค Bagging และ Boosting ไม่ช่วยเพิ่มประสิทธิภาพการจำแนกข้อมูล หรือกล่าวอีกนัยหนึ่งคือ ถ้าต้องการใช้เทคนิค Multiple learning จะต้องไม่ให้ข้อมูลในคลาสใดคลาสหนึ่ง มีปริมาณสูงเกินกว่าคลาสอื่นๆ ที่เหลืออย่างมีนัยสำคัญ

5.2.3 การลดคุณลักษณะด้วยวิธี Correlation-based Feature Subset Selection (CFS) และวิธี Feature selection method based on correlation measure and relevance & redundancy analysis (FCBF) ส่งผลให้ประสิทธิภาพความถูกต้อง (Accuracy) ในการจำแนก

ประเภทของโรคใกล้เคียงกับการไม่ลดคุณลักษณะ แต่สามารถลดมิติของข้อมูลลงได้อย่างมาก ซึ่งการลดมิติของข้อมูลดังกล่าว ทำให้ประหยัดทรัพยากรของระบบคอมพิวเตอร์ในการเรียนรู้เพื่อสร้างโมเดล ทั้งหน่วยประมวลผลและหน่วยความจำได้เป็นอย่างมาก เมื่อนำไปพัฒนาเป็นซอฟต์แวร์วินิจฉัยโรคอัตโนมัติ ย่อมส่งผลให้ระบบซอฟต์แวร์โดยรวมมีประสิทธิภาพดีขึ้นอย่างมีนัยสำคัญ

### 5.3 ข้อเสนอนแนะ

จากการทดลองการจำแนกกลุ่มข้อมูลทางการแพทย์ เพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ โดยมีจุดมุ่งหมายที่จะทดสอบประสิทธิภาพของอัลกอริทึมต่างๆ ที่ใช้ในการทำเหมืองข้อมูล มุ่งเน้นค้นหาอัลกอริทึมที่เหมาะสมที่สุดสำหรับฐานข้อมูลทางการแพทย์ พบว่าไม่มีอัลกอริทึมใดที่สามารถทำงานได้ดีที่สุดกับทุกกลุ่มข้อมูล สามารถสรุปได้ว่าลักษณะของข้อมูลแต่ละประเภทส่งผลต่อประสิทธิภาพการทำงานของแต่ละอัลกอริทึม มีข้อเสนอแนะว่าควรศึกษาพัฒนาอัลกอริทึมที่สามารถใช้ได้กับทุกกลุ่มข้อมูลทางการแพทย์ โดยให้ประสิทธิภาพออกมาดีกับทุกลักษณะข้อมูล และจากการทดลอง เวลาในการสังเคราะห์โมเดลจากข้อมูลทั้ง 13 ชุด พบว่าบางอัลกอริทึม ใช้เวลาในการสังเคราะห์โมเดลนานมาก รวมถึงใช้หน่วยประมวลผลและหน่วยความจำปริมาณมาก โดยเฉพาะอัลกอริทึม Multilayer Perceptron, Radial Basis Function Network ดังนั้น ควรมีการศึกษา วิธีการลดระยะเวลา และทรัพยากรของระบบในการประมวลผลเพื่อการสังเคราะห์โมเดล รวมถึงควรมีการศึกษาวิธีการลดคุณลักษณะแบบอื่นๆ ที่แตกต่างจากงานวิจัยดังกล่าว ที่ส่งผลให้มิติของข้อมูลลดลงและเพิ่มประสิทธิภาพในการจำแนกกลุ่มข้อมูลทางการแพทย์ให้สูงขึ้น