

การจำแนกข้อมูลการเกิดคดีอาญาโดยใช้เทคนิคเหมืองข้อมูลกรณีศึกษา เขตอำเภอศรีราชา จังหวัดชลบุรี

The classification techniques of criminal case data. In Sriracha Amphur, Chonburi province

วิศรุท อีสริยศไกร, กสิณภาส ทันจันทิก และ สุพาพร บรรดาศักดิ์

ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา

Emails: visarut.i@ku.th, kasinpat.t@ku.th, jumbundasak@hotmail.com

บทคัดย่อ

การศึกษานี้เป็นการวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความเสี่ยงในการเกิดคดีอาญา ในพื้นที่เขตอำเภอศรีราชา จังหวัดชลบุรี โดยได้เก็บรวบรวมข้อมูลจากการขอข้อมูลคดีจากสถานีตำรวจภูธรศรีราชา อำเภอศรีราชา จังหวัดชลบุรี เพื่อทำเป็นเหมืองข้อมูล (Data Mining) โดยใช้หลักการพยากรณ์ (Prediction) แล้วนำไปจำแนกข้อมูล (Classification) โดยนำข้อมูลไปกำหนดจุดเสี่ยงการเกิดคดีอาญาต่างๆ โดยแบ่งระดับความเสี่ยงแยกออกเป็นสี่ ขึ้นแสดงบนแผนที่ (Google Maps API) โดยการบอกจุดที่อยู่ ณ ขณะนั้น เพื่อบอกความเสี่ยง และสามารถแนะนำวิธีการลดความเสี่ยงได้ และมีระบบประเมินการบ่งชี้ความเป็นผู้ร้าย โดยการวิเคราะห์ออกมาเป็น Web Application ซึ่งจะครอบคลุมเขตพื้นที่อำเภอศรีราชา จังหวัดชลบุรี

คำสำคัญ-- เหมืองข้อมูล; การพยากรณ์; คดีอาญา;
จำแนกข้อมูล

ABSTRACT

This study analyzes the factors that affect the risk of criminal prosecution. Area Sriracha Chonburi province the data collected from the information from the police station in Sriracha, Sriracha, Chonburi. To make data mining (Data Mining) using the prediction

(Prediction) then classified information. (Classification) by bringing data to determine the risk of various criminal cases. Divided by risk level color split. Showing up on a map (Google Maps API) by saying that at that point to say the risk. And to recommend ways to reduce the risk. And an evaluation system to identify the culprit. The analysis comes as the Web Application, which cover areas Sriracha. Chonburi province

Keywords-- classification; data mining;
predictive; criminal case

1.บทนำ

คดีอาญา เป็นคดีที่ผู้กระทำความผิดมีเจตนาในการกระทำความผิด มีลักษณะที่ร้ายแรงเป็นอันตรายต่อสังคม ซึ่งก่อให้เกิดความเดือดร้อนต่อสังคม ทั้งความเสียหาย ต่อ ทรัพย์สิน หรือ บุคคล ตัวอย่างเช่น คดีทำร้ายร่างกาย คดีลักทรัพย์ คดีชิงทรัพย์ คดีปล้นทรัพย์ คดีฆ่าคนตาย คดีประมาททำให้ผู้อื่นบาดเจ็บ หรือ เสียชีวิต คดีข่มขืน คดีรับของโจร เป็นต้น จากข้อมูลของสำนักงานเทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานตำรวจแห่งชาติ พบว่าได้มีผู้เสียหายจากคดีอาญาในรูปแบบต่างๆ จำนวนมาก โดยผู้เสียหายส่วนใหญ่จะเป็นเพศหญิง เช่น คดีปล้นทรัพย์ คดีชิงทรัพย์ คดีข่มขืน เป็นต้น ในส่วนเพศชายจะเป็นคดีทำร้ายร่างกาย หรือ คดีฆ่าคนตาย โดยปัจจัยเสี่ยงในการเกิด

คดีอาญาเหล่านี้มีได้หลายสาเหตุตั้งแต่ เวลา สถานที่ หรือแม้แต่ เพศซึ่งส่งผลต่อการเกิดคดีอาญาทั้งสิ้น

ในการเก็บข้อมูลเพื่อจะนำมาวิเคราะห์ด้วยเทคนิคเหมือง ข้อมูล จะใช้วิธีการนำข้อมูลจากสถานีตำรวจภูธรศรีราชา อำเภอ ศรีราชา จังหวัดชลบุรี มาใช้ในเทคนิคเหมืองข้อมูล โดยในระบบ การจำแนกข้อมูลการเกิดคดีอาญา เป็นระบบที่จะแสดงผลการ วิเคราะห์ความเสี่ยงการเกิดคดีอาญาออกมาให้ผู้ใช้งานได้ทราบ และระบบจะช่วยบอกวิธีการป้องกัน หรือหลีกเลี่ยง เพื่อลดความ เสี่ยงในการเกิดคดีอาญา ซึ่งจะสามารถช่วยลดอัตราผู้เสียหาย จากคดีอาญาให้ลดน้อยลง ทั้งความเสียหายทางร่างกาย ความ เสียหายทางทรัพย์สิน และความเสียหายทางจิตใจให้ลดน้อยลง ได้

2. ทฤษฎีและการวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำ กับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อน อยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไป ประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการ ตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้ง ในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการ หนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บ ข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถ ดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถ ค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทาง การเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์ เพื่อเอาข้อมูล ขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ ทันสมัย

- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่ หรือฐานข้อมูล

- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทาง สถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่ สามารถช่วยการตัดสินใจได้

ขั้นตอนการทำเหมืองข้อมูล

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่ เกี่ยวข้องออกไป

- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลาย แหล่งให้เป็นข้อมูลชุดเดียวกัน

- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการ วิเคราะห์จากแหล่งที่บันทึกไว้

- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้ เหมาะสมสำหรับการใช้งาน

- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์ จากข้อมูลที่มีอยู่

- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้ จากการทำเหมืองข้อมูล

- Knowledge Representation เป็นขั้นตอนการนำเสนอ ความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

2.2 Ripper Algorithm

ขั้นตอนวิธีริปปเปอร์ (Ripper Algorithm) เป็นการเรียนรู้กฎ ซึ่งมีกระบวนการตัดแต่งกิ่งที่ เพิ่มขึ้นซ้ำๆ เพื่อลดข้อผิดพลาดที่ เกิดจากการสร้างกฎ ขึ้น (Cohen, 1995) มี 4 ขั้นตอน ขั้นตอนที่ 1 เป็นการเจริญเติบโต (Growth) สร้าง ลำดับของกฎเฉพาะโดย การเพิ่มกฎจนกระทั่งได้กฎเป็น ที่น่าพอใจแล้วจึงจะหยุดเพิ่ม ขั้นตอนที่ 2 เป็นการตัดแต่งกิ่ง (Pruning) โดยตัดกฎที่ลด ประสิทธิภาพการทำงานของการเรียนรู้กฎออก ขั้นตอนที่ 3 การ เพิ่มประสิทธิภาพ (Optimization) มีการเพิ่มคุณลักษณะเข้าไป ในแต่ละกฎเดิมหรือกฎที่ถูก สร้างขึ้นใหม่ในขั้นตอนที่ 1 และ 2 ขั้นตอนที่ 4 การเลือกกฎ (Selection) เป็นการ เลือกกฎที่ดีที่สุด เก็บไว้ ส่วนกฎอื่นๆ จะถูกตัดออกไป โดยค่าความยาวของ คำอธิบายที่น้อยที่สุด (Minimum Description Length)

2.3 ต้นไม้ตัดสินใจ (Decision Tree)

เป็นการทำงานของต้นไม้ตัดสินใจโดยที่จะทำการแบ่ง Note Attribute ของข้อมูล โดยที่จะมี leafed Node เป็นผลลัพธ์ของ ข้อมูล ต้นไม้ตัดสินใจมีข้อดี คือ แปลความหมายได้ง่าย โดยจะทำการแปลงต้นไม้ตัดสินใจ ให้เป็นกฎได้เลย

2.4 การเรียนรู้แบบเบย์ (Naïve Bayes)

การเรียนรู้แบบเบย์เป็นหนึ่งในเทคนิคของ Classification โดยจะใช้หลักความน่าจะเป็นที่เรียกว่า Conditional probability โดยจะทำตามเงื่อนไข เพื่อหาผลลัพธ์ของข้อมูล

2.5 การวิจัยที่เกี่ยวข้อง

พรพิรุณ โอ่งอินทร์ (2558) นำเสนอข้อมูลการวิเคราะห์ปัจจัยที่มีความสัมพันธ์ต่อการเกิดอาชญากรรมด้วยวิธีการทำเหมืองข้อมูล จังหวัดขอนแก่น โดยใช้วิธีการที่มีอยู่ใน Weka (JRip) หรือ อัลกอริทึม Ripper ซึ่งเป็นการเรียนรู้เงื่อนไขที่เพิ่มขึ้นเรื่อยๆ ซ้ำๆ เพื่อให้มีความผิดพลาดน้อยลง จากกฎที่สร้างขึ้น โดยใช้ข้อมูลคดีอาญาทั้งหมด 5 ประเภทเพื่อนำมาวิเคราะห์เพื่อให้เจ้าหน้าที่ตำรวจได้ดูแลความสงบ ได้อย่างสะดวกสบายมากยิ่งขึ้น โดยการศึกษาความสัมพันธ์จะศึกษาการเกิดเหตุอาชญากรรมแล้วนำมาวิเคราะห์ ด้วยวิธีการเก็บข้อมูลโดยเทคนิค Data mining เพื่อทำการวิเคราะห์ข้อมูลการเกิดคดีได้อย่างมีประสิทธิภาพสูงสุด

ปฐมพงษ์ ฉับพลัน และ ฐิมาพร เพชรแก้ว (2553) นำเสนอการประยุกต์ใช้ Google maps API ในการเฝ้าระวังปัญหาอาชญากรรม โดยเป็นระบบเฝ้าระวังเพื่อแก้ไขปัญหาอาชญากรรมโดยใช้ Google API โดยจะเฝ้าระวังเป็นพื้นที่แหล่งเสี่ยงของผู้ค้า ผู้เสพ สถานที่ค้า และจุดตรวจ โดยมีการแสดงผลบน Google map เพื่อให้ง่ายต่อการทราบของตำรวจ ช่วยในการเฝ้าระวังหรือจับกุม ได้อย่างมีประสิทธิภาพ สามารถแสดงสถิติการจับกุมคดีที่เกี่ยวข้องกับยาเสพติด โดยจะมีลักษณะเป็นกราฟในแต่ละเดือน เพื่อแสดงปัญหาอาชญากรรมแต่ละประเภทได้อย่างชัดเจน โดยใช้โปรแกรมที่มีการพัฒนาแบบ Web-based application สามารถทำงานผ่านระบบอินเทอร์เน็ตได้ โดย Google maps API เป็นสิ่งที่ใช้ในการเชื่อมต่อกับ Google เพื่อใช้ดูข้อมูลทางภูมิศาสตร์ที่ Google มี แล้วสร้างขอบเขตบนแผนที่โดยใช้ GPolyline() และ สร้าง ตัวควบคุมบนแผนที่โดยใช้ GMapTypeControl() ของ Google maps API

อรรถพงษ์ ศรีเหนียง (2557) นำเสนอการสำรวจพื้นที่ในชุมชน โดยจะตรวจสอบและเก็บข้อมูลทางภูมิศาสตร์โดยใช้ Google maps เพื่อกำหนดเป้าหมายในการสำรวจ โดยเมื่อได้เป้าหมาย ก็จะทำการปักหมุดและจัดเก็บข้อมูลลงใน Database แล้วนำข้อมูลมาวิเคราะห์ติดตามโรคเพื่อนำมาสำรวจ ช่วยให้การ

ปัญหาการเข้าถึงข้อมูลของการเดินสำรวจ และยังสามารถติดตามพฤติกรรมของผู้ที่อาศัยอยู่ภายในชุมชนได้อีกด้วย โดยจะเป็นเว็บไซต์ ซึ่งเว็บไซต์จะพัฒนาโดยใช้ภาษา C# ทำงานร่วมกับ SQL Server และมีการใช้ .NET Framework และ ASP .NET ในส่วนแผนที่จะใช้ Google maps API ดึงเข้าไปในเว็บไซต์ที่สร้างโดยเขียนเป็น HTML , Javascript ในแบบที่ไม่ซับซ้อน ปัญหาที่พบอาจเกิดการปัดหมุดคลาดเคลื่อนได้

มณฑล เขี่ยมไพศาล และมานัส ศรีวนิช (2553) นำเสนอข้อมูลเรื่องความหนาแน่นเชิงพื้นที่อาชญากรรม กรณีศึกษาเขตพระนคร กรุงเทพมหานคร ศึกษาตามแบบเคอร์เนล โดยใช้แผนที่อธิบายการเกิดคดีอาชญากรรมโดยเทคนิคเชิงพื้นที่ โดยจะมีเทคนิคร่วมในการใช้ Dot map เป็นการแสดงข้อมูลเป็นจุดบนแผนที่ โดยจะแบ่งออกมาเป็นกรณีได้อีก 3 ประเภท ได้แก่ 1. Central tendency เป็นการวัดหาตำแหน่งกลาง 2. Clustering and Dispersion เป็นการวัดการรวมตัวและการกระจาย 3. Distance from a point เป็นการวัดระยะทางจากจุดใดจุดหนึ่ง และมีเทคนิคการวัดการรวมตัวได้เป็นกรณีได้อีก 3 กรณี ได้แก่ 1) มัธยฐานของจุด หรือ การกระจายจากตำแหน่งตัวกลาง 2) Specific location เป็นการกระจายจากตำแหน่งที่ตั้งเฉพาะ 3) การกระจายโดยพิจารณาความสัมพันธ์จุดอื่นข้างเคียง โดยเป็นการวัดการกระจายที่เป็นแบบจุด

ปราโมทย์ ครอบงูญ (2556) นำเสนอข้อมูลการวิเคราะห์เชิงสถิติหลายตัวแปรของคดีอาญาระดับจังหวัดของประเทศไทย ซึ่งเป็นการวิเคราะห์ข้อมูลคดีระดับจังหวัดโดยใช้เทคนิค 3 ตัวแปร ได้แก่ 1. วิเคราะห์ปัจจัย 2. วิเคราะห์การจัดกลุ่ม 3. วิเคราะห์ Discriminant ซึ่งมีวิธีการวิจัยดังนี้ โดยเริ่มจาก Factor เป็นการหาปัจจัย ของคดีเพื่อให้ตัวแปรเป็นตัวแปรที่มีมาตรฐาน ก่อนที่จะนำไปวิเคราะห์ จากนั้นเมื่อได้ปัจจัยจากการวิเคราะห์เรียบร้อยแล้ว ก็จะมาทำการวิเคราะห์โดยการจัดกลุ่ม เพื่อให้ได้กลุ่มของแต่ละจังหวัดที่มีการเกิดคดีที่คล้ายกันให้อยู่ในกลุ่มเดียวกัน แล้วจึงจะทำการวิเคราะห์ด้วย Discriminant

3. วัตถุประสงค์ของเนื้อเรื่อง

3.1 เพื่อศึกษาลักษณะความสัมพันธ์การเกิดคดีอาญา จากการเก็บรวบรวมข้อมูล จากปัจจัยเสี่ยง

3.2 เพื่อให้ได้ระบบวิเคราะห์ความเสี่ยงในการเกิดคดีอาญา

3.3 เพื่อให้ผู้ใช้งานระบบรู้ถึงระดับความเสี่ยงในการเกิดคดีอาญาต่อตนเอง

3.4 เพื่อลดความเสี่ยงในการเกิดคดีอาญาให้แก่ผู้ใช้ระบบ

4. วิธีการดำเนินการ

การดำเนินการวิจัยมีทั้งหมด 5 ขั้นตอน ดังต่อไปนี้

4.1 Business Understanding เป็นการทำความเข้าใจในงานที่จะทำ โดยในขั้นตอนนี้จะหาปัญหาที่ทำให้เกิดคดีและหาแนวทางแก้ไขต่างๆ โดยใช้หลักการการทำเหมืองข้อมูลโดยปัญหาที่พบคือในเขตอำเภอศรีราชามีอันตรายการเกิดคดีที่ค่อนข้างสูง จึงได้นำข้อมูลการเกิดคดีในพื้นที่เขตอำเภอศรีราชา มาทำการจำแนกข้อมูล (Classification) เพื่อจำแนกประเภทคดีที่มีความเสี่ยงในระดับต่างๆ โดย Target Class เป็นระดับความเสี่ยงที่เกิดในแต่ละพื้นที่ เพื่อนำมาทำนาย และลดปัญหาการเกิดคดี

4.2 Data Understanding เป็นการทำความเข้าใจในข้อมูล โดยที่นำข้อมูลบันทึกประจำวันคดีที่เกิดขึ้นในแต่ละวัน มาทำการจำแนกข้อมูล (Classification) เพื่อจำแนกประเภทคดีที่มีความเสี่ยงในระดับต่างๆ โดยมีชุดข้อมูลฝึกสอน (Training set) ดังตารางที่ 1 ประกอบด้วยคุณลักษณะ (Attribute) ทั้งหมด 29 คุณลักษณะ โดยมี Target class คือ Danger_Level เป็นระดับความเสี่ยงที่เกิดในแต่ละพื้นที่ เพื่อนำมาทำนาย และลดปัญหาการเกิดคดี ประกอบด้วยคุณสมบัติต่อไปนี้

ตาราง 1. คุณลักษณะที่ใช้ในการทำนาย

คุณลักษณะ	คำอธิบาย
Type_Day	ประเภทวันที่เกิดคดีวันธรรมดาและวันหยุด (weekdays, holiday)
Time_case	เวลาการเกิดคดี (เช่น 17.01-18.00 เป็นต้น)
Time_At_night	เวลายามวิกาล (เช่น หลัง 18.00 เป็น night , ก่อน 18.00 เป็น day)
Sex_Offender	เพศผู้กระทำความผิด (female,male)
Age_Offender	อายุผู้กระทำความผิด (เช่น 26-30 ปี)
location_case	สถานที่เกิดคดี (เป็นข้อมูลที่เจาะจง เช่น ในบ้านเลขที่ 3 ม.1 ต.ศรีราชา)
location_type	ชนิดสถานที่เกิดคดี (เช่น ในหมู่บ้าน)

Sexual_age	ช่วงวัยของผู้กระทำความผิด (แบ่งตามช่วงวัยเช่น Child อายุต่ำกว่า 25)
Deck_cards	มีไพ่หรือไม่ (Yes,No)
Bet_Gamblers	มีการวางเดิมพันหรือไม่ (Yes,No)
Chit_lottery	มีโพยหวยหรือไม่ (Yes,No)
Drunk_Offender	ผู้กระทำความผิดเมาหรือไม่ (Yes,No)
Drive_Offender	ผู้กระทำความผิดได้ขับรถหรือไม่ (Yes,No)
Hit_Item	มีการทำลายข้าวของหรือไม่ (Yes,No)
Mayhem_Offender	ผู้กระทำความผิดทำร้ายร่างกายหรือไม่ (Yes,No)
Kill_human	ผู้กระทำความผิดฆ่าคนหรือไม่ (Yes,No)
Weapon	ผู้กระทำความผิดใช้อาวุธหรือไม่ (Yes,No)
Theft_Item	ผู้กระทำความผิดขโมยของหรือไม่ (Yes,No)
Theft_Human	ผู้กระทำความผิดลักพาตัวเด็กหรือคนหรือไม่ (Yes,No)
Amphetamine	ผู้กระทำความผิดยุ่งเกี่ยวกับยาเสพติดหรือไม่ (Yes,No)
Privacy	ละเมิดความเป็นส่วนตัว (Yes,No)
Rape	มีข่มขืนผู้เสียหายหรือไม่ (Yes,No)
Act_Check	ทำผิด พรบ.เช็คหรือไม่ (Yes,No)
Act_Animal	ทำผิด พรบ.สัตว์หรือไม่ (Yes,No)
Act_Shop	ทำผิด พรบ.เปิดร้าน,การขายของหรือไม่ (Yes,No)
Act_Car	ทำผิด พรบ.ความเร็วหรือผิดเกี่ยวกับรถ (Yes,No)
Act_Warrior	ทำผิด พรบ.ทหาร (Yes,No)
Act_Computer	ทำผิด พรบ.คอมพิวเตอร์ (Yes,No)
Act_foreign_work	ทำผิด พรบ.แรงงานต่างชาตินหรือไม่ (Yes,No)
Danger_Level	ระดับความอันตราย แบ่งเป็น 5 ระดับ (Very Low, Low, Medium, Dangerous, Very Dangerous)

4.3 Data Preparation เป็นขั้นตอนการเตรียมข้อมูลก่อนนำไปสร้างโมเดล โดยการนำข้อมูลบันทึกประจำวันที่ได้มาทำการ cleaning ข้อมูลเพื่อให้มีความถูกต้อง ซึ่งบางข้อมูลอาจเกิดการผิดพลาด เช่นมีอายุที่สูญหาย จึงทำการแก้ไขด้วยการหาค่ากลางของอายุโดยการนำอายุที่มีทั้งหมดมาค่าเฉลี่ยแล้วนำมาใส่ในข้อมูลอายุที่หายไป หรือข้อมูลอายุที่เกินไปจากความเป็นจริง เช่น อายุของคนทั่วไปนั้น จะมีอายุไม่เกิน 90 ปี แต่มีบางคนอายุ 500 ปี จึงต้องทำการแก้ไขข้อมูลนี้เนื่องจากไม่ถูกต้องเพราะอาจเกิดการพิมพ์ข้อมูลผิดพลาด จึงแก้ไขด้วยวิธีการลบ 0 ออกไปตัวหนึ่งเพื่อไม่ให้อายุเกินจากคนทั่วไป เป็นต้น

4.4 Model Building เป็นการนำข้อมูลที่มีความถูกต้องจากการเตรียมข้อมูลมาสร้างโมเดล โดยการหาอัลกอริทึมที่เหมาะสม เพื่อให้ได้โมเดลที่เหมาะสมกับข้อมูล จากการทดสอบใช้อัลกอริทึม 3 อัลกอริทึม

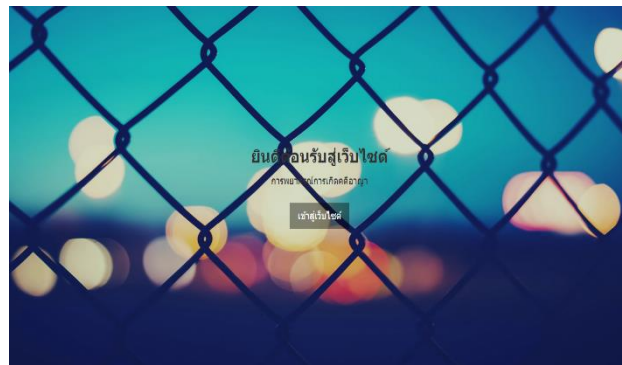
- 1) Ripper Algorithm (JRip) ผลที่ได้มีค่าความถูกต้อง 94.2272 %
- 2) Decision Tree (J48) ผลที่ได้มีค่าความถูกต้อง 95.1583 %
- 3) Naïve Bayes (Naïve Bayes) ผลที่ได้มีค่าความถูกต้อง 91.0615 %

4.5 Testing and Evaluation จากการสร้างโมเดลที่ได้จาก 3 อัลกอริทึม

- 1) Ripper Algorithm (JRip)
- 2) Decision Tree (J48)
- 3) Naïve Bayes (Naïve Bayes)

จะพบว่า Decision Tree (J48) มีค่าความถูกต้องมากที่สุดดังนั้นจึงได้โมเดลที่เหมาะสมกับข้อมูลมากที่สุด

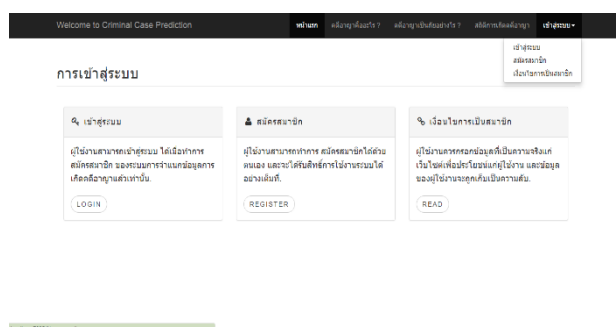
4.6 Web Application เมื่อได้โมเดลที่ผ่านขั้นตอนการทำเหมืองข้อมูลเรียบร้อยแล้ว ก็จะนำโมเดลที่ได้มาไว้ใน Web Application เพื่อทำการวิเคราะห์ความเสี่ยงในการเกิดคดีอาญาในเขตอำเภอศรีราชา จังหวัดชลบุรี ดังตัวอย่างต่อไปนี้



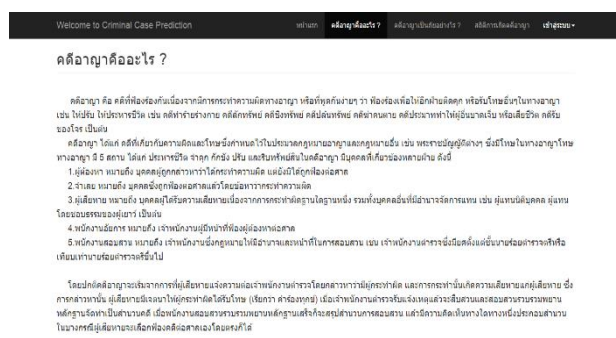
รูปที่ 1. หน้าแรกเมื่อเข้าเว็บไซต์



รูปที่ 2. หน้าแสดงการต้อนรับสู่เว็บไซต์



รูปที่ 3. หน้าแสดงการเข้าสู่ระบบสำหรับสมาชิก และสมัครสมาชิก



รูปที่ 4. หน้าแสดงรายละเอียดคดีอาญา



รูปที่ 5. หน้าแสดงรายละเอียดคดีส่งผลกระทบอย่างไร



รูปที่ 6. หน้าสถิติการเกิดคดี

โดยรูปที่ 1 – 6 จะเป็นรูปผู้ใช้งานทั่วไปสามารถเข้ามาใช้งานได้ สามารถดูรายละเอียดต่างๆเกี่ยวกับคดีอาญาว่าคืออะไร ส่งผลต่อตนเองอย่างไร และสถิติที่เกิดคดีอาญาเป็นอย่างไร เมื่อผู้ใช้งานทั่วไปสนใจ ที่จะวิเคราะห์ว่าตนเองมีความเสี่ยงแค่ไหนกับการเกิดคดีอาญา ก็สามารถสมัครสมาชิก แล้วทำการล็อกอินเข้าสู่ระบบเพื่อวิเคราะห์ความเสี่ยงในการเกิดคดีอาญากับตนเองได้ ดังรูปต่อไปนี้



รูปที่ 7. หน้าการล็อกอินสำหรับสมาชิกเว็บไซต์

เมื่อผู้ใช้งานที่เป็นสมาชิกต้องการวิเคราะห์ว่าตนเองมีความเสี่ยงกับการเกิดคดีอาญาหรือไม่ ก็จะมีรายละเอียดต่างๆ ให้กรอก ผู้ใช้งานก็จะทำการกรอก เมื่อกดวิเคราะห์ระบบก็จะทำการวิเคราะห์ ผู้ใช้งานเพียงรอสักครู่ ระบบก็จะแจ้งผลการวิเคราะห์ออกมา เป็นระดับความอันตราย และบอกพื้นที่เสี่ยงที่เป็นอันตรายนั้น

การวิเคราะห์

อายุ

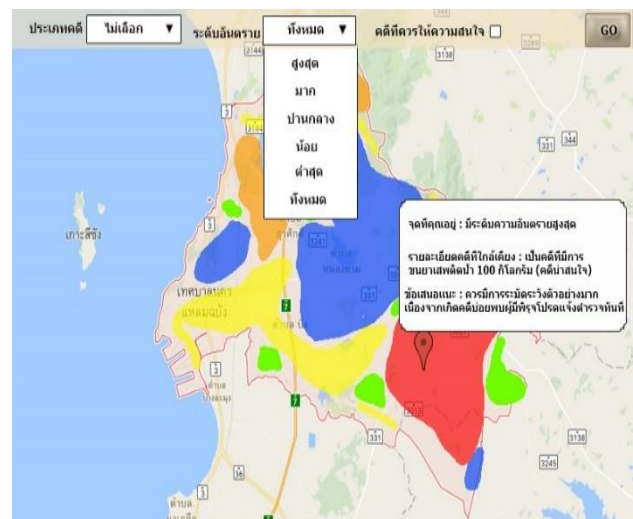
เพศ

อายุ

ชื่อที่อยู่

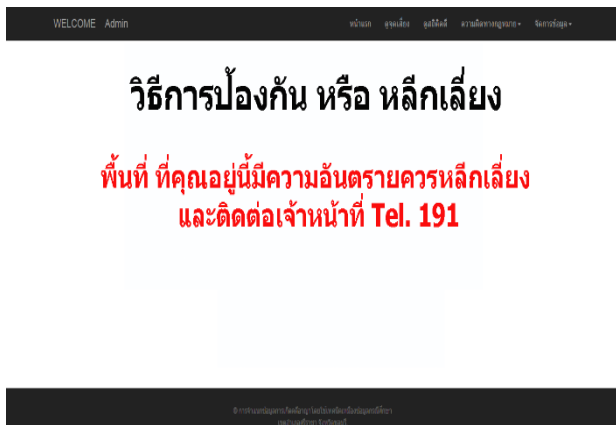
ประเภทที่อยู่

รูปที่ 8. หน้าวิเคราะห์สำหรับสมาชิกเท่านั้น



รูปที่ 9. หน้าผลการวิเคราะห์แสดงออกมาบน Google maps

เมื่อแสดงผลวิเคราะห์ออกมาเรียบร้อยแล้ว ระบบจะมีการ
แนะนำวิธีการป้องกัน หรือหลีกเลี่ยง ให้กับผู้ใช้งานอีกด้วย เพื่อ
ลดความเสี่ยงในการเกิดคดีอาญาให้น้อยลง



รูปที่ 10. หน้าวิธีการป้องกัน

5. ผลการทดลอง

จากผลการทดลองเมื่อนำข้อมูลไปวิเคราะห์ด้วยโปรแกรม Weka โดยใช้ 3 อัลกอริทึม ดังต่อไปนี้ Ripper Algorithm(JRip), Decision Tree(J48) และ Naïve Bayes(Naïve Bayes) จึงได้ว่า

JRip มีความถูกต้อง 94.2272 %

=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
144	4	1	0	0	a = Medium
2	218	0	0	8	b = Very Low
0	0	18	2	0	c = Very Dangerous
0	2	0	66	4	d = Dangerous
2	5	0	1	60	e = Low

รูปที่ 11. Confusion Matrix ของ JRip

J48 มีความถูกต้อง 95.1583 %

=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
145	3	1	0	0	a = Medium
0	220	0	0	8	b = Very Low
0	0	18	2	0	c = Very Dangerous
0	1	0	67	4	d = Dangerous
4	2	0	1	61	e = Low

รูปที่ 12. Confusion Matrix ของ J48

Naïve Bayes มีความถูกต้อง 91.0615 %

=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
135	3	3	3	5	a = Medium
5	213	0	2	8	b = Very Low
2	1	13	4	0	c = Very Dangerous
4	1	1	64	2	d = Dangerous
1	3	0	0	64	e = Low

รูปที่ 13. Confusion Matrix ของ Naïve Bayes

ผลการทดลองเมื่อนำ 3 อัลกอริทึมมาเปรียบเทียบกัน



รูปที่ 14. สรุปผลการทดลอง

จากภาพที่ 14 จะเห็นได้ว่าผลการทดลองโดยการใช้ อัลกอริทึม J48 มีความถูกต้องสูงที่สุด โดย JRip และ Naïve Bayes มีความถูกต้องรองลงมาตามลำดับ โดยที่ J48 มีความถูกต้องคือ 95.1583%

6. สรุปและอภิปรายผล

เนื้อหาเป็นการทำเหมืองข้อมูล (Data Mining) โดยใช้ข้อมูลจากสถานีตำรวจภูธรศรีราชา อำเภอศรีราชา จังหวัดชลบุรี ซึ่งนำข้อมูลที่ได้มาผ่านขั้นตอนการทำเหมืองข้อมูล (Data Mining) โดยใช้หลักการจำแนกข้อมูล (Classification) โดยมี 3 อัลกอริทึมหลักที่ใช้ คือ Ripper Algorithm , Decision Tree , Naïve Bayes โดยผลลัพธ์ที่ได้คือการทำนายความเสี่ยงในการเกิดคดี ซึ่งสามารถแบ่งเป็นระดับความเสี่ยงได้ 5 ระดับ ดังนี้

- 1) สีน้ำเงิน คือ ความอันตรายต่ำสุด
- 2) สีเขียว คือ ความอันตรายน้อย
- 3) สีเหลือง คือ ความอันตรายปานกลาง
- 4) สีส้ม คือ ความอันตรายมาก
- 5) สีแดง คือ ความอันตรายสูงสุด

ผลลัพธ์ในการสร้างโมเดลของแต่ละอัลกอริทึม คือ

1) Ripper Algorithm (JRip) ผลที่ได้มีค่าความถูกต้อง
94.2272 %

2) Decision Tree (J48) ผลที่ได้มีค่าความถูกต้อง
95.1583 %

3) Naïve Bayes (Naïve Bayes) ผลที่ได้มีค่าความ
ถูกต้อง 91.0615 %

จึงสรุปได้ว่าอัลกอริทึมในการสร้างโมเดลที่เหมาะสมกับข้อมูลชุด
นี้ที่สุด คือ Decision Tree (J48) ที่ความถูกต้อง 94.2272 %
ดังนั้นโมเดลที่ได้นี้ จึงมีความเหมาะสมในการใช้ทำนายความ
เสี่ยงการเกิดคดี

เอกสารอ้างอิง

- [1] มณฑล เขียมไพศาล, มานัส ศรีวณิช. “ความหนาแน่น
เชิงพื้นที่อาชญากรรม กรณีศึกษาเขตพระนคร
กรุงเทพมหานคร ศึกษาตามแบบเคอร์เนล”.
[ออนไลน์] 2553 [อ้างเมื่อ 21 กันยายน 2559]. จาก
<http://www.tds.tu.ac.th/jars/download/jars/v7-1/07%20Criminal%20Pattern.pdf>
- [2] ปฐมพงษ์ ฉับพลัน, จิมาพร เพชรแก้ว. “การ
ประยุกต์ใช้ Google Maps API ในการเฝ้าระวังปัญหา
ยาเสพติด”. [ออนไลน์] 2553 [อ้างเมื่อ 23 กันยายน
2559]. จาก
http://lexitron.nectec.or.th/public/NCIT_2010_Bangkok%20_Thailand/index_files/papers/37-p095.pdf
- [3] พรพิรุณ โองอินทร์. “การวิเคราะห์ปัจจัยที่มี
ความสัมพันธ์ต่อการเกิดอาชญากรรมด้วยวิธีการทำ
เหมืองข้อมูล”. [ออนไลน์] 2558 [อ้างเมื่อ 23
กันยายน 2559]. จาก
<https://gsbooks.gs.kku.ac.th/58/the34th/pdf/P MO3.pdf>
- [4] ปราโมทย์ ครอบงูญ. “การวิเคราะห์เชิงสถิติหลายตัว
แปรของคดีอาญาระดับจังหวัดของประเทศไทย”.
[ออนไลน์] 2556 [อ้างเมื่อ 23 กันยายน 2559]. จาก

http://www.resjournal.kku.ac.th/abstract/18_4_9.pdf

- [5] อรรถพงษ์ ศรีเหนียง. “โปรแกรมประยุกต์บนเว็บเพื่อ
การสำรวจชุมชนสำหรับคณะพยาบาลศาสตร์
มหาวิทยาลัยสยาม”. [ออนไลน์] 2557 [อ้างเมื่อ 23
กันยายน 2559]. จาก <http://www.research-system.siam.edu/2013-12-20-03-59-31/2013-12-20-04-08-38/1645-2013-12-20-05-58-136>
- [6] Sajee Vanich. “Data Mining (เหมืองข้อมูล)”.
[ออนไลน์] 2558 [อ้างเมื่อ 18 กันยายน 2559]. จาก
<http://sajeegm301.blogspot.com/>
- [7] OK Nation Blog. “กระบวนการยุติธรรมทางอาญา”.
[ออนไลน์] 2550 [อ้างเมื่อ 18 กันยายน 2559]. จาก
<http://www.oknation.net/blog/korung/2009/12/12/entry-1>