

การพยากรณ์รูปแบบของการกระทำความผิดซ้ำของเด็กและเยาวชน

The forecasting patterns of juvenile recidivism

ศิริลักษณ์ หล่อพันธุ์มณี*, ศิริพร ฉิมพลี, วิชชา ฉิมพลี, วัชรารักษ์ เนตรหาญ และณรงค์ฤทธิ์ ภิรมย์นก

สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต

บทคัดย่อ

การกระทำความผิดของเด็กและเยาวชน เป็นประเด็นปัญหาที่หลายประเทศต่างให้ความสำคัญเป็นอันดับต้น ๆ เนื่องจากเด็กและเยาวชนจะเป็นกำลังที่สำคัญของชาติในอนาคตและการกระทำความผิดซ้ำของเด็กและเยาวชนมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ซึ่งการวิจัยนี้ได้ทำการศึกษารูปแบบของการกระทำความผิดของเด็กและเยาวชน เพื่อคาดการณ์การกระทำความผิดซ้ำของเด็กและเยาวชน โดยการประยุกต์ใช้อัลกอริธึมปัญญาประดิษฐ์ (AI : Artificial Intelligence) ผลการวิจัยพบว่าการพยากรณ์ที่ใช้ปัจจัยที่เป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชนและถูกคัดเลือกโดยอัลกอริธึม XGBoost ประกอบด้วย 6 คุณลักษณะ 1) การดำรงชีพด้วยการประกอบอาชีพ 2) พฤติกรรมจัดแจงเข้าบ้าน อาคาร หรือรถของคนอื่น 3) การรับฝาก รับจ้างซื้อและหรือจำหน่ายยาเสพติด 4) ประวัติการเคยถูกดำเนินคดี 5) ประวัติการฝึกอบรม/จำคุก/คุมประพฤติ/เข้ารับการฟื้นฟูสมรรถภาพผู้ติดยาเสพติด 6) อายุขณะกระทำความผิด ซึ่งผลการประเมินประสิทธิภาพของอัลกอริธึมมีดังนี้ ค่าความถูกต้อง คิดเป็นร้อยละ 97.80 ค่าความแม่นยำ คิดเป็นร้อยละ 97.80 และค่าความระลึก คิดเป็นร้อยละ 100

คำสำคัญ: กระทำความผิดซ้ำของเด็กและเยาวชน ปัญญาประดิษฐ์ อัลกอริธึม XGBoost ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก

Abstract

The juvenile delinquency is one of the most important problems of many countries because juvenile will be important forces of the nation in the future and the recidivism of juvenile tend to increase continuously. The research also studies the forms of juvenile delinquency to predict juvenile recidivism by applying the AI: Artificial Intelligence algorithm. The result shows the predictions based on factors that cause juvenile delinquency and selected by the XGBoost algorithm consists of 6 characteristics: 1) Living by making a living 2) Tamper behavior into someone else's house, building, or car 3) Accepting employment, purchasing and / or selling drugs 4) History of prosecution 5) Training history / Imprisonment / Probation / Rehabilitation of drug addicts 6) Age

while committing crime. The results of the performance evaluation are as follows. 97.80 percent of the accuracy accounted for 97.80 percent and the gift accounted for 100 percent.

Keywords: the juvenile recidivism, Artificial Intelligence, XGBoost algorithm, Accuracy, Precision, Recall

บทนำ

การกระทำความผิดของเด็กและเยาวชน เป็นประเด็นปัญหาที่หลายประเทศต่างให้ความสำคัญเป็นอันดับต้น ๆ ในการกำหนดนโยบายเพื่อการแก้ไขอย่างเร่งด่วน เนื่องจากเด็กและเยาวชนจะเป็นกำลังที่สำคัญของชาติในอนาคต แต่การดำเนินการของกระบวนการยุติธรรมที่เน้นการคุมครองเด็กและเยาวชนมากเกินไป ก็อาจทำให้การแก้ไขปัญหาไม่สามารถดำเนินการได้อย่างเต็มประสิทธิภาพเท่าที่ควร และอาจเกิดปัญหาเกี่ยวกับการกระทำความผิดซ้ำของเด็กและเยาวชนตามมาอย่างหลีกเลี่ยงไม่ได้ เนื่องจากการไม่เกรงกลัวต่อบทลงโทษหรือผลกระทบที่จะได้รับการกระทำความผิดถึงแม้ว่ารัฐบาลของแต่ละประเทศจะมีความพยายามที่จะพัฒนาและสนับสนุนให้เด็กและเยาวชนเป็นทรัพยากรมนุษย์ที่มีคุณภาพ แต่สิ่งที่สะท้อนให้เห็นในสถานการณ์ปัจจุบัน พบว่า มีเด็กและเยาวชนจำนวนไม่น้อยที่มีพฤติกรรมไม่เหมาะสม ตั้งแต่การฝ่าฝืนกฎระเบียบของโรงเรียน การมั่วสุม การรวมกลุ่มเพื่อทะเลาะวิวาท รวมถึงการสร้างควมรุนแรงที่เพิ่มขึ้นจนนำไปสู่การกระทำความผิดกฎหมาย เป็นต้น ซึ่งรูปแบบการกระทำความผิดของเด็กและเยาวชนในปัจจุบันนับได้ว่ามีความรุนแรงและมีลักษณะเช่นกับการกระทำความผิดของผู้ใหญ่มากยิ่งขึ้น

ข้อมูลจากกรมพินิจและคุ้มครองเด็กและเยาวชนในปี พ.ศ. 2555 – 2558 (กรมพินิจและคุ้มครองเด็กและเยาวชน, 2562) พบว่าการกระทำความผิดซ้ำของเด็กและเยาวชนในพื้นที่กรุงเทพมหานคร มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ถึงแม้ที่ผ่านมาจะมีการศึกษาวิจัยเพื่อค้นหาแนวทางการแก้ไขปัญหาที่เกิดขึ้น แต่ยังพบว่าสถานการณ์ที่เด็กและเยาวชนมีการกระทำความผิดซ้ำมิได้ลดลงแต่อย่างใด โดยในปี พ.ศ. 2558 มีจำนวนคดีที่เด็กและเยาวชนกระทำความผิดซ้ำเพิ่มขึ้น 2,177 คดี คิดเป็นร้อยละ 52.76 ซึ่งพื้นที่ในกรุงเทพมหานครมีความหลากหลายของลักษณะทางประชากร จึงทำให้ประสบปัญหาความรุนแรงของอาชญากรรมอยู่ในระดับที่สูงกว่าในเขตพื้นที่จังหวัดอื่น ๆ โดยการศึกษาวิจัยเรื่องแผนที่ความรุนแรงของอาชญากรรมในประเทศไทย ของปรารถนา อุดพัฒนา ปี พ.ศ. 2560 พบว่า กรุงเทพมหานครเป็นพื้นที่ที่มีความรุนแรงของอาชญากรรมมากที่สุด รองลงมาเป็นจังหวัดนครราชสีมา ชลบุรี เชียงใหม่ ร้อยเอ็ด สมุทรปราการ ปทุมธานี สงขลา อุบลราชธานี สุราษฎร์ธานี และนครศรีธรรมราช ตามลำดับ โดยผู้วิจัยได้อธิบายว่าสาเหตุที่พื้นที่ดังกล่าวมีอาชญากรรมสูง เนื่องจากมีประชากรอยู่หนาแน่น มีการกระจุกตัวของอุตสาหกรรม และมีการขยายตัวทางเศรษฐกิจจำนวนมาก (ปรารถนา อุดพัฒนา, 2560)

ตลอดระยะเวลาที่ผ่านมา การแก้ไขปัญหาอาชญากรรมในพื้นที่กรุงเทพมหานคร ได้มีการพัฒนากระบวนการดำเนินงานโดยใช้เทคโนโลยีสารสนเทศมาเป็นเครื่องมือในการวิเคราะห์เพิ่มมากขึ้น เช่น มีการใช้แผนที่อาชญากรรม (Crime Mapping) ในการนำข้อมูลทางภูมิศาสตร์และข้อมูลอาชญากรรมมาทำการวิเคราะห์พื้นที่เสี่ยง เช่น สถานที่เกิดเหตุอาชญากรรม วันเวลาที่เกิดเหตุ ความถี่ ความต่อเนื่อง หรือแนวโน้มของการเกิดอาชญากรรม เป็นต้น ซึ่งการวิเคราะห์แผนที่อาชญากรรมนั้น มีการใช้ข้อมูลจากเทคโนโลยีสารสนเทศทางภูมิศาสตร์ (Geographic Information System) เป็นตัวสำคัญในการดำเนินการ อย่างไรก็ตาม การใช้เทคนิคดังกล่าวยังเป็นการมุ่งเน้นที่การจัดการกับอาชญากรรมโดยทั่วไปในกลุ่มผู้ใหญ่นั้น ยังไม่ได้มีการพัฒนาเทคโนโลยีทางสารสนเทศภูมิศาสตร์สำหรับการป้องกันการกระทำความผิดของเด็กและเยาวชนที่เป็นการเฉพาะ จึงทำให้การป้องกันการกระทำความผิดของเด็กและเยาวชนต้อง

ใช้แนวทางเช่นเดียวกับที่ดำเนินการกับผู้กระทำผิดที่เป็นผู้ใหญ่ ทั้งที่มีบริบท รายละเอียดของการกระทำความผิด และวิธีการแก้ปัญหาที่แตกต่างกัน ดังนั้นจึงมีความจำเป็นที่จะต้องมีการศึกษาและวิจัยสำหรับการวิเคราะห์ปัจจัยเพื่อหาสาเหตุของการกระทำผิด รวมถึงเทคโนโลยีการพยากรณ์แนวโน้มการกระทำผิดซ้ำของเด็กและเยาวชนในแต่ละพื้นที่ เพื่อประโยชน์ต่อการป้องกันการกระทำความผิดของเด็กและเยาวชนในแต่ละพื้นที่ของกรุงเทพมหานคร รวมถึงการจัดทำมาตรการสำหรับการกำกับติดตามไม่ให้เด็กและเยาวชนที่ได้รับการปล่อยตัวคืนสู่สังคมหวนกลับมากระทำผิดซ้ำอีก ซึ่งจะเป็นการตอบโต้ภัยการวิจัยเพื่อป้องกันการและแก้ไขปัญหาเด็กและเยาวชนที่กระทำความผิดที่สอดคล้องกับสังคมในยุคดิจิทัล และสนับสนุนการดำเนินงานเพื่อแก้ไขปัญหาเร่งด่วนของการอำนวยความสะดวกยุติธรรมและการบริหารงานยุติธรรมสำหรับเด็กและเยาวชนตามนโยบายไทยแลนด์ 4.0 ต่อไป

วัตถุประสงค์

- 1) เพื่อศึกษาและจัดทำข้อมูลด้านรูปแบบและปัจจัยที่เป็นสาเหตุการกระทำความผิดของกลุ่มเด็กและเยาวชนในพื้นที่กรุงเทพมหานคร
- 2) เพื่อพัฒนาระบบฐานข้อมูลเด็กและเยาวชนที่กระทำความผิดและได้รับการปล่อยตัวคืนสู่สังคมในพื้นที่กรุงเทพมหานคร สำหรับการจัดทำมาตรการกำกับติดตามไม่ให้หวนกลับมากระทำผิดซ้ำ
- 3) เพื่อสร้างระบบสารสนเทศทางภูมิศาสตร์สำหรับการป้องกันการกระทำความผิดของกลุ่มเด็กและเยาวชนในแต่ละพื้นที่ในเขตกรุงเทพมหานคร

วิธีการวิจัย

2.1 แนวคิดและทฤษฎีที่เกี่ยวกับการกระทำความผิดของเด็กและเยาวชน

การกระทำความผิดของเด็กและเยาวชนมาจากปัจจัยหลายประการ แต่ละบุคคลก็อาจมีสาเหตุที่แตกต่างกัน เด็กและเยาวชนบางคนอาจพบว่าปัญหามาจากสาเหตุเดียว แต่บางคนอาจมาจากหลายปัจจัยรวมกัน ดังนั้นการเข้าใจความหมายของการกระทำความผิดของเด็กและเยาวชนจึงเป็นสิ่งสำคัญ

วรชาติ แส่นคำ ได้ให้ความหมายของเด็กและเยาวชนที่กระทำความผิด หมายถึง เด็กและเยาวชนที่กระทำความผิดและถูกตำรวจจับดำเนินคดีในข้อหาต่าง ๆ ซึ่งข้อหาเหล่านั้นเป็นความผิดตามประมวลกฎหมายอาญาและพระราชบัญญัติที่มีโทษทางอาญา (วรชาติ แส่นคำ อ้างถึงใน ชื่นจิตร วาณิชยสุวรรณ, 2554, น.7)

ประเทือง ธนียผล ได้สรุปการกระทำความผิดของเด็กและเยาวชนว่าเป็นการกระทำความผิดกฎหมายอาญามาตรา 11 และมีความผิดที่เกี่ยวข้องกับพฤติกรรมไม่เหมาะสมของเด็กและเยาวชน ได้แก่ มีพฤติกรรมเกรี้ยวกราดรุนแรงหนีโรงเรียน (ประเทือง ธนียผล อ้างถึงใน ชื่นจิตร วาณิชยสุวรรณ, 2554, น.6)

จากที่กล่าวมาข้างต้นการกระทำความผิดของเด็กและเยาวชนเป็นความผิดทางอาญาที่ต้องมีบุคคลที่ทำหน้าเข้าไปช่วยเหลือ ดูแล และฟื้นฟูจิตใจของเด็กและเยาวชนที่กระทำความผิด เพื่อให้เด็กและเยาวชนที่กระทำความผิดนั้นสามารถกลับเข้าสู่สังคมและไม่กระทำความผิดซ้ำอีก

การกระทำผิดซ้ำของเด็กและเยาวชนเป็นเรื่องที่มีผู้ให้ความสนใจมาก เพราะเด็กและเยาวชนอยู่ในช่วงอายุระหว่างวัยเด็กและวัยรุ่นตอนต้น ซึ่งเป็นช่วงวัยที่พบว่ามีปัญหามากที่สุด โดยเฉพาะในช่วงอายุ 13-18 ปี ดังนั้นกลุ่มเด็กและเยาวชนควรได้รับการดูแลและป้องกัน เพื่อพัฒนาเด็กและเยาวชนเหล่านั้นให้เป็นกำลังของประเทศต่อไป (ชาญคณิต กฤตยา สุริยะมณี และอุนิษา เลิศโตมรสกุล, 2553, น.1)

สาเหตุของการกระทำความผิดของเด็กและเยาวชนเกิดจากหลาย ๆ ปัจจัยมาประกอบกัน สามารถสรุปได้ออกเป็น 3 ด้าน คือ

1) ความไม่รู้หรือเฝ้าระวังของเด็กและเยาวชน ทำให้รู้เท่าไม่ถึงการณ์ จึงมักตกเป็นเครื่องมือให้บุคคลอื่นชักจูงไปกระทำความผิดได้ง่าย

2) การได้เห็นพฤติกรรมที่ไม่ดีจากบุคคลในครอบครัว การเลี้ยงดูจากครอบครัวที่มีบุคคลในครอบครัวกระทำความผิดมาก่อน เด็กและเยาวชนซึ่งไม่มีความหนักแน่นจะถูกชักจูงได้ง่าย ให้มีพฤติกรรมลอกเลียนแบบและกระทำความผิด (โสภา ขปิลมันน์, 2536, น.131-132) ได้ศึกษาปัจจัยทางด้านครอบครัวที่มีผลต่อพฤติกรรมเด็กและเยาวชน พบว่าสาเหตุของการกระทำความผิดที่เป็นผลมาจากครอบครัวอาจมีสาเหตุ ดังนี้

(1) ความสัมพันธ์ระหว่างสมาชิกในครอบครัวและภาวะความเป็นอยู่ เด็กและเยาวชนที่กระทำความผิดมาจากครอบครัวที่แตกแยก บิดาและมารดาหย่าร้างกัน แยกกันอยู่ บิดา/มารดาถึงแก่กรรม รวมทั้งความไม่สงบสุขของครอบครัวที่มีสาเหตุมาจากบิดาและมารดาทะเลาะเบาะแว้ง ทุกทีกัน เด็กและเยาวชนที่ต้องอยู่ในสภาพแวดล้อมดังกล่าวจะขาดความมั่นคงทางสภาพจิตใจและอารมณ์ บางครั้งอาจต้องตกเป็นเครื่องมือของบิดาและมารดา ทำให้เด็กและเยาวชนเสียใจ ผิดหวัง และแสดงออกโดยการกระทำความผิด (จรรยา สุวรรณทัต อ้างถึงใน ชื่นจิตร์ วาณิชสุพรรณ, 2554, น.7)

(2) ลักษณะการอบรมเลี้ยงดูจากบิดา มารดา หรือผู้ปกครองที่มีความเข้มงวดมากจนเป็นเผด็จการหรือในทางตรงข้ามไม่สนใจดูแลเด็กและเยาวชน เมื่อมีการกระทำความผิดแล้วลงโทษเด็กและเยาวชนด้วยความรุนแรงเกินไป

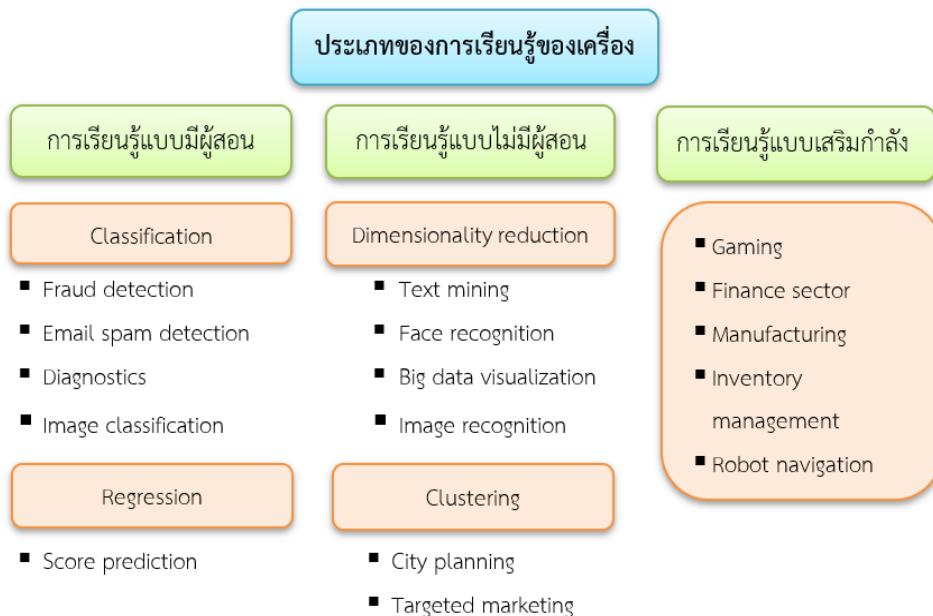
(3) ปัจจัยด้านเศรษฐกิจ ครอบครัวมีฐานะยากจน ความเป็นอยู่ย่ำแย่ จึงเป็นสาเหตุบีบบังคับให้เด็กและเยาวชนกระทำความผิด (อนันต์ ประวิเศษ, 2547, น.14) ปัญหาจากการว่างงาน สมาชิกในครอบครัวไม่มีงานทำ ทำให้ไม่มีรายได้จึงเป็นปัจจัยทำให้เด็กและเยาวชนไปกระทำความผิด (นวลจันทร์ ทศนชัยกุล, 2548, น.109)

3) ความผิดปกติของร่างกายและทางอารมณ์ ทำให้เด็กและเยาวชนขาดสติความยับยั้งชั่งใจ หรือไม่สามารถควบคุมตนเองได้ จึงมีพฤติกรรมต่อต้านและกระทำความผิดได้

4) ปัจจัยด้านสังคมและสิ่งแวดล้อม (สุชา จันทรธม, 2541, น.23-38) ได้สรุปปัจจัยด้านสังคมและสิ่งแวดล้อมที่เป็นสาเหตุให้เด็กและเยาวชนกระทำความผิดมาจากการเพิ่มจำนวนของประชากร ชุมชนและสภาพความเป็นอยู่ไม่ดี เด็กและเยาวชนได้เห็นตัวอย่างที่ไม่ดีจากบุคคลในชุมชน ทำให้เกิดเป็นความเคยชินต่อการเห็นการกระทำความผิด ทำให้เด็กและเยาวชนกระทำความผิดเลียนแบบผู้ใหญ่ (โกศล วงศ์สวรรค์ และคณะ, 2537, น.251)

2.2 การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine learning) เป็นวิธีการที่ทำให้คอมพิวเตอร์สามารถเรียนรู้ได้โดยอัตโนมัติ พร้อมทั้งจดจำวิธีการตัดสินใจเพื่อแก้ปัญหาที่เกิดขึ้น ๆ แต่ไม่มีกฎเกณฑ์ที่ชัดเจนในการตัดสินใจ (Zhang & TSAI, 2003) การพัฒนาระบบการเรียนรู้ของเครื่องประกอบด้วย 2 ขั้นตอนขั้นตอนแรก คือ การเรียนรู้โดยการนำเอาข้อมูลเก่ามาเรียนรู้ (Training) เพื่อให้คอมพิวเตอร์สามารถเรียนรู้และจดจำรูปแบบ (Pattern) เพื่อนำไปสู่การสร้างตัวแบบ (Model) สำหรับการพยากรณ์หรือการตัดสินใจในการแก้ไขปัญหา ขั้นตอนที่สองเป็นการพยากรณ์หรือการตัดสินใจในการแก้ไขปัญหา (Prediction) โดยใช้ข้อมูลในปัจจุบันที่ไม่เคยเรียนรู้มาก่อน นำเข้าไปที่ตัวแบบที่ได้สร้างไว้ เพื่อทำการวิเคราะห์และตัดสินใจในการแก้ไขปัญหา



ภาพที่ 1 ประเภทของการเรียนรู้ของเครื่อง

ประเภทของการเรียนรู้ของเครื่อง แบ่งออกได้ 3 ประเภท ดังภาพที่ 1 มีรายละเอียด ดังนี้

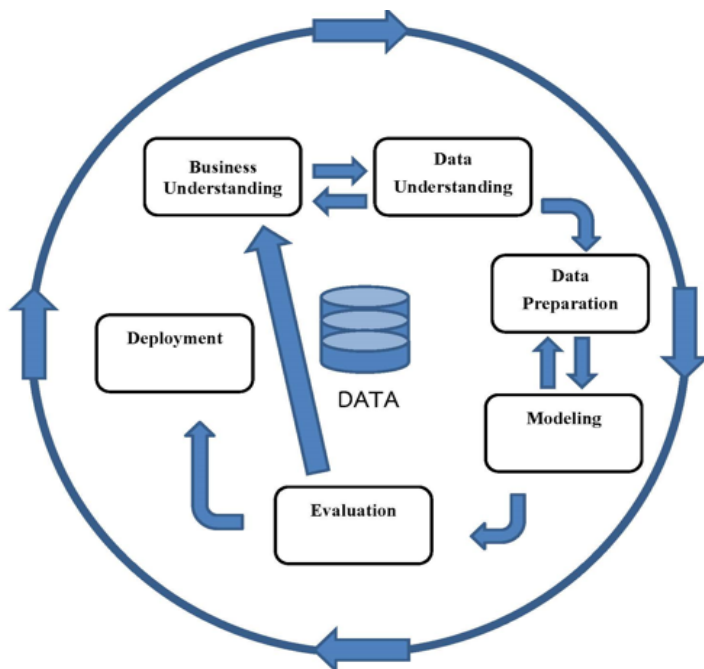
1) การเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นการเรียนรู้โดยใช้ข้อมูลในอดีตที่มีทั้งข้อมูลนำเข้า (Input) และผลลัพธ์ (Output) เพื่อให้อัลกอริธึมสร้างตัวแบบที่เกิดจากการเรียนรู้ระหว่างข้อมูลนำเข้าและผลลัพธ์ โดยมีความคลาดเคลื่อน (Error) น้อยที่สุด อัลกอริธึมที่อยู่ในการเรียนรู้ประเภทนี้ ได้แก่ การจำแนกประเภท (Classification) การถดถอย (Regression) เป็นต้น

2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เป็นการเรียนรู้โดยใช้ข้อมูลนำเข้า (Input) แต่ไม่ทราบผลลัพธ์ (Output) การหาผลลัพธ์ว่าควรเป็นอย่างไร ต้องอาศัยวิธีการหรืออัลกอริธึมอื่นมาช่วยในการสร้างการเรียนรู้ อัลกอริธึมที่อยู่ในการเรียนรู้ประเภทนี้ ได้แก่ การจัดกลุ่ม (Clustering) การลดจำนวนมิติเพื่อบีบอัดข้อมูล (Dimensionality reduction) การวิเคราะห์โครงข่าย (Social network analysis) เป็นต้น

3) การเรียนรู้แบบเสริมกำลัง (Reinforcement learning) เป็นการเรียนรู้ที่คอมพิวเตอร์จะสนใจต่อสิ่งแวดล้อมรอบตัวเป็นพิเศษ เช่น ในการเล่นเกมกรุก เกมส OX เป็นต้น ผู้เล่นเกมกรุกต้องคาดการณ์ล่วงหน้าว่าจะเกิดเหตุการณ์อะไรขึ้นได้ในการเดินหมากรุกแต่ละครั้ง ซึ่งการเดินแต่ละครั้งอาจจะไม่เป็นผลดีต่อครั้งนั้น แต่อาจมีผลดีต่อการเดินหมากรุกในครั้งถัดไปได้ อัลกอริธึมที่อยู่ในการเรียนรู้ประเภทนี้ ได้แก่ Q-learning, Deep Q-Networks (DQN) และ State-Action-Reward-State-Action (SARSA) เป็นต้น

ในขั้นตอนของการเรียนรู้ของเครื่อง งานวิจัยนี้ใช้กระบวนการมาตรฐานในการวิเคราะห์ข้อมูล เรียกว่า กระบวนการมาตรฐานอุตสาหกรรมข้ามสำหรับการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining: CRISP-DM) ที่มีกระบวนการจัดการข้อมูลที่ต้อง สามารถจัดกลุ่ม ค้นหาความสัมพันธ์ของกลุ่มข้อมูล และนำมาสร้างแบบจำลองเพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคตได้ (Shearer, 2000) เพื่อให้ระบบมีการเรียนรู้แบบอัตโนมัติ

กระบวนการ CRISP-DM ประกอบด้วย 6 ขั้นตอน ดังภาพที่ 2 แต่ละขั้นตอนมีการทำงานที่ต่อเนื่องกัน โดยขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้า ซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกล่องสี่เหลี่ยมแต่ละกล่อง ตัวอย่างเช่นเมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูล (Data Preparation) แล้วจะนำไปสร้างโมเดลจำแนกประเภทข้อมูลในขั้นตอน Modeling และหลังจากนั้นอาจจะย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้น เพื่อหวังว่าจะได้แบบจำลองที่มีความถูกต้องมากขึ้น ขั้นตอนในกระบวนการ CRISP-DM มีรายละเอียด ดังนี้



ภาพที่ 2 กระบวนการมาตรฐานอุตสาหกรรมข้ามสำหรับการทำเหมืองข้อมูล (Shearer, 2000)

1) Business Understanding

ขั้นตอนแรกของกระบวนการที่ต้องทำความเข้าใจกับปัญหาและแปลงปัญหาให้เป็นโจทย์สำหรับนำไปใช้ในการวิเคราะห์ข้อมูล โดยการสัมภาษณ์และรับฟังปัญหาหรือความต้องการจากผู้บริหาร เจ้าหน้าที่กรมพินิจและคุ้มครองเด็กและเยาวชน ที่จะเป็นผู้นำผลการวิเคราะห์ข้อมูลไปใช้ประโยชน์ โดยความต้องการทั้งหมดจะนำมาจัดลำดับความสำคัญ และกำหนดวัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูลของหน่วยงาน เช่น ผู้บริหารกรมพินิจและคุ้มครองเด็กและเยาวชน ต้องการทราบปัจจัยสำคัญที่ทำให้เด็กและเยาวชนกระทำความผิดซ้ำ จำนวนของเด็กและเยาวชนที่กระทำความผิด 3 ปีซ้อนหลัง จำแนกตามเขตของกรุงเทพมหานคร สัดส่วนของจำนวนเด็กและเยาวชนที่กระทำความผิดต่อจำนวนเด็กและเยาวชนทั้งหมดจำแนกตามเขตของกรุงเทพมหานคร เป็นต้น

2) Data Understanding

ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นทำการตรวจสอบข้อมูล เพื่อดูความถูกต้องของข้อมูล สำหรับการสร้างฐานข้อมูลให้ครบถ้วน ตลอดจนตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยการพิจารณาว่าจะใช้

ข้อมูลทั้งหมดหรือคัดเลือกเฉพาะข้อมูลที่สำคัญมาใช้ในการวิเคราะห์ซึ่งต้องมีความสอดคล้องกับวัตถุประสงค์ที่กำหนดไว้ งานวิจัยนี้ได้เก็บรวบรวมข้อมูลจากกรณีศึกษาและคุ้มครองเด็กและเยาวชนที่ได้ทำการสัมภาษณ์เด็กและเยาวชนที่กระทำผิดตามแบบสัมภาษณ์เพื่อประเมินความเสี่ยงและความจำเป็น และข้อมูลการกระทำผิดของเด็กและเยาวชนที่ได้จากการบันทึกคดีของสถานีตำรวจ ที่แสดงให้เห็นข้อมูลการกระทำผิดซ้ำของเด็กและเยาวชนหลังได้รับการปล่อยตัว

3) Data Preparation

ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมาให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ หรือที่เรียกว่าข้อมูลพร้อมใช้ โดยในขั้นตอนนี้ต้องทำการจัดเตรียมข้อมูลให้ถูกต้อง เช่น การแปลงข้อมูลให้เหมาะสมสำหรับการวิเคราะห์ การเติมข้อมูลที่ขาดหายไป เป็นต้น

4) Modeling

ขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล เช่น การจำแนกประเภทข้อมูล การแบ่งกลุ่มข้อมูล การพยากรณ์ข้อมูล เป็นต้น งานวิจัยได้ใช้วิธีการจำแนกประเภทด้วยอัลกอริธึม XGBoost สำหรับการจำแนกเด็กและเยาวชนที่กระทำผิดและกระทำผิดซ้ำ มีรายละเอียดดังนี้

(1) การจำแนกประเภทของข้อมูล

การจำแนกประเภทเป็นเทคนิคที่ใช้ในการจำแนกคลาสของข้อมูลด้วยคุณลักษณะต่าง ๆ ที่ได้มีการกำหนดไว้แล้ว โดยแบ่งข้อมูลออกเป็น 2 ส่วน ข้อมูลส่วนที่ 1 ใช้สำหรับให้ระบบได้เรียนรู้และจำแนกข้อมูลส่วนที่ 1 ออกมาเป็นประเภทตามที่ได้กำหนดไว้ ผลที่ได้จากการเรียนรู้ คือ ตัวจำแนกประเภทของข้อมูล ข้อมูลส่วนที่ 2 จะถือว่าเป็นข้อมูลใหม่ที่ยังไม่เคยเรียนรู้มาก่อน ตัวจำแนกประเภทของข้อมูลจะทำหน้าที่จัดประเภทและทำนายว่าข้อมูลส่วนที่สองนี้จะจัดอยู่ในประเภทใด

ภาพที่ 3 แสดงรายละเอียดของขั้นตอนการทำงานของการสร้างแบบจำลองจำแนกประเภทข้อมูล การจำแนกประเภทข้อมูลจะเกิดขึ้นจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลเรียนรู้หรือข้อมูลที่ใช้สำหรับสอน (Training set) เป็นชุดข้อมูลที่มีบทบาทในการสร้างแบบจำลองข้อมูลขึ้นมา และกลุ่มข้อมูลทดสอบ (Test set) เป็นชุดข้อมูลที่ใช้ประเมินความถูกต้องของแบบจำลองที่สร้างขึ้น กระบวนการสร้างแบบจำลองจำแนกประเภท ประกอบด้วย 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การสร้างแบบจำลอง

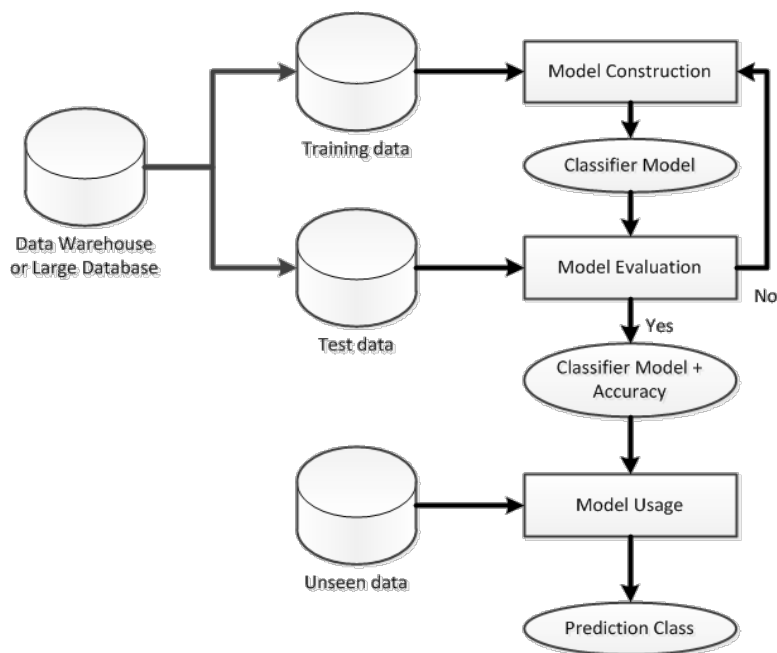
การสร้างแบบจำลอง (Model construction) โดยใช้ข้อมูลที่ได้เรียนรู้เสร็จเรียบร้อยแล้ว ซึ่งแต่ละตัวจะมีคุณลักษณะหนึ่งที่บอกว่าข้อมูลอยู่ในประเภทใด ข้อมูลนี้จะถูกเรียกว่า ข้อมูลสอน (Training data) แบบจำลองที่ได้มาจากการใช้อัลกอริธึมใดอัลกอริธึมหนึ่งที่ตั้งอยู่ในการจำแนกประเภทของข้อมูล เช่น ถ้าใช้อัลกอริธึมต้นไม้ตัดสินใจ (Decision tree) แบบจำลองที่ได้จะมีลักษณะคล้ายต้นไม้จริงแต่กลับหัว มีโหนดรากอยู่ด้านบนสุดและโหนดใบอยู่ล่างสุดของต้นไม้ แต่ละโหนดบนต้นไม้จะมีคุณลักษณะ (Attribute) เป็นตัวเลือกทดสอบ มีกิ่งที่เป็นค่าที่เป็นไปได้ของคุณลักษณะ (Attribute value) ที่ถูกเลือกทดสอบและมีโหนดใบแสดงคลาสที่กำหนดไว้

ขั้นตอนที่ 2 การทดสอบความแม่นยำของแบบจำลอง

แบบจำลองที่ได้จากการเรียนรู้ต้องถูกตรวจสอบความถูกต้อง (Evaluation of model accuracy) โดยใช้ข้อมูลที่ยังไม่ได้เรียนรู้มาก่อนสำหรับทดสอบเรียกว่า ข้อมูลทดสอบ (Test data) ข้อมูลทดสอบเป็นชุดข้อมูลที่ใช้ในการเรียนรู้ กลุ่มที่ถูกต้องของข้อมูลทดสอบจะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากแบบจำลองจำแนกประเภทเพื่อประเมินว่าแบบจำลองที่ได้นี้สามารถจัดกลุ่มประเภทข้อมูลได้อย่างถูกต้องมากน้อยเพียงใด

ขั้นตอนที่ 3 การนำแบบจำลองที่ได้ไปใช้

การนำแบบจำลองที่ได้ไปใช้ (Model usage) กับข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen data) เพื่อทำนายและกำหนดประเภทให้กับข้อมูลนั้น ข้อมูลทดสอบเป็นอิสระไม่ขึ้นกับข้อมูลสอน ผลที่ได้จะถูกนำไปหาค่าความแม่นยำ (Accuracy) ของแบบจำลองที่ได้ก่อนนำไปใช้



ที่มา: ชินพัฒน์ แก้วชินพร, (2553)

(2) การพยากรณ์ข้อมูล

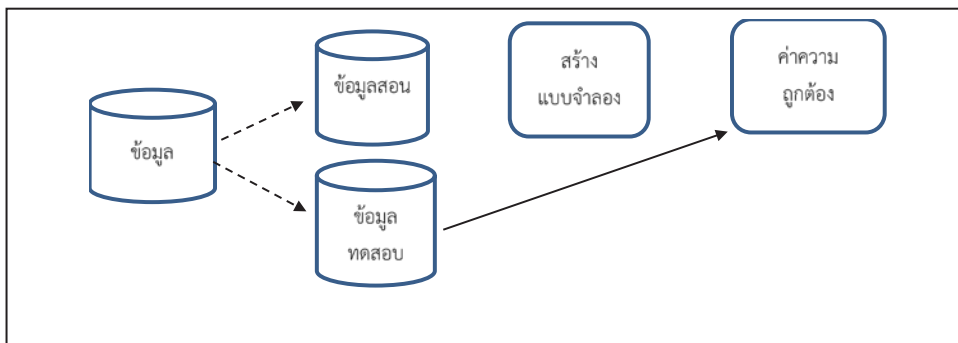
การพยากรณ์ข้อมูล (Data prediction) เป็นกระบวนการสร้างแบบจำลองเพื่อพยากรณ์ค่าของข้อมูลที่ต้องการทราบจากข้อมูลที่มีอยู่ การพยากรณ์ข้อมูลไม่มีการจัดข้อมูลเข้ากลุ่มตามที่ได้กำหนด แต่เป็นการพยากรณ์หาค่าที่ต้องการออกมาเป็นตัวเลข ตัวอย่างเช่น การพยากรณ์จำนวนเด็กและเยาวชนที่จะกระทำผิดซ้ำ เป็นต้น

นงเยาว์ ในอรุณ (2557: 34) อธิบายกระบวนการพยากรณ์ข้อมูล ดังภาพที่ 4 ประกอบด้วย

(2.1) การเรียนรู้เพื่อสร้างแบบจำลอง โดยการเรียนรู้จากข้อมูลสอน แบบจำลองที่ได้สามารถเป็นไปได้หลายรูปแบบ

(2.2) การประมาณความถูกต้องของแบบจำลองจากข้อมูลทดสอบ ค่าที่ได้นี้จะนำไปเปรียบเทียบกับค่าที่ได้จากแบบจำลอง เพื่อประเมินค่าความถูกต้อง

(2.3) การนำแบบจำลองไปใช้ เป็นการนำแบบจำลองมาใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน โดยการกำหนดค่าผลลัพธ์ให้กับข้อมูลที่ไม่เคยเห็นมาก่อนนี้



ที่มา: Han, Kamber and Pei (2011: 364)

(3) Extreme Gradient Boosting

Extreme Gradient Boosting หรือ XGBoost มีพื้นฐานมาจาก Gradient Boosting ถูกจัดอยู่ในกลุ่มของการจำแนกประเภทแบบ Boosting ประเภทหนึ่ง ที่ได้รับการพัฒนาให้มีประสิทธิภาพมากขึ้น จึงทำให้สามารถทำงานได้รวดเร็วและมีประสิทธิภาพมากขึ้น สามารถจัดการ หน่วยการทำงานย่อยแบบ Multithread ได้ดีและการทำให้ตัวแปรมีค่าอยู่ในระดับปกติ (Regularization) เป็นการป้องกันการเข้ากันได้มากเกินไป (Over fit) ผลการเรียนรู้จะได้เป็นแบบจำลองที่นำเอาต้นไม้ตัดสินใจมาสร้างให้เกิดการเรียนรู้แบบต่อเนื่องต่อกัน การเรียนรู้จากการสร้างต้นไม้ตัดสินใจหลาย ๆ ต้น จะทำให้เกิดการเรียนรู้โดยสนใจจากค่าความคลาดเคลื่อน หรือเกิดการเรียนรู้จากความคลาดเคลื่อนของต้นไม้ตัดสินใจที่สร้างได้ก่อนหน้า ทำให้ความแม่นยำของการพยากรณ์แม่นยำมากขึ้นเรื่อย ๆ เมื่อการเรียนรู้ของต้นไม้ตัดสินใจดำเนินอย่างต่อเนื่องจนมีความลึกมากพอหรือไม่มีรูปแบบอื่น ๆ ให้เรียนรู้เพิ่ม ระบบจะหยุดการเรียนรู้ แบบจำลองที่ได้จากอัลกอริธึม XGBoost จะใช้แบบจำลองหลาย ๆ ตัวแบบมาประกอบกันเป็นแบบจำลองที่ซับซ้อนขึ้น ดังเช่นในงานวิจัยของ Chen & Guestrin. (2016) ที่นำไปใช้ในการพยากรณ์ผู้ต้องสงสัยที่ไม่ประสงค์ดีต่อระบบ

(4) Evaluation

ขั้นตอนการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรกหรือไม่ และมีความน่าเชื่อถือมากน้อยเพียงใด ก่อนที่จะนำผลมาวิเคราะห์ไปใช้งาน หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า สำหรับงานวิจัยนี้ใช้วิธีการประเมินประสิทธิภาพด้วยวิธีการทดสอบไขว้ แบบแบ่งเป็น 5 ส่วน (5-folds cross validation)

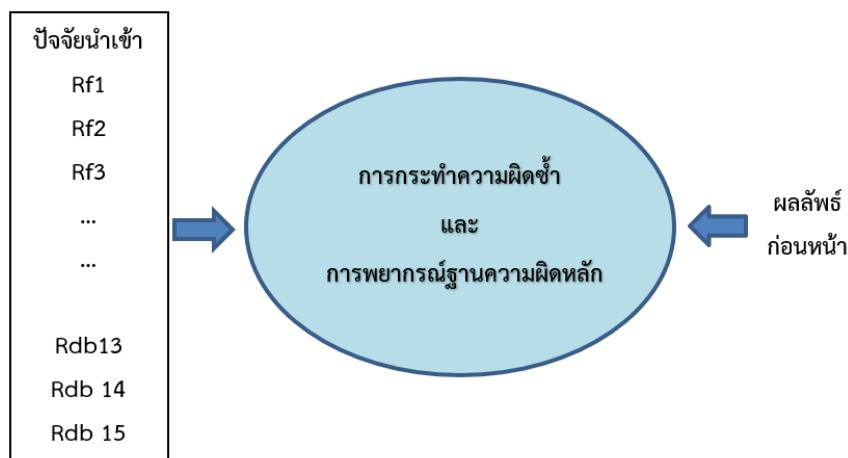
(5) Deployment

การนำผลลัพธ์ที่ได้ไปใช้ในสถานการณ์จริงเป็นการนำผลการวิจัยไปให้หน่วยงานที่เกี่ยวข้องใช้งานจริงติดตาม และประเมินผลการใช้งาน เพื่อนำกลับไปแก้ไขและปรับปรุงให้ตรงกับความต้องการของผู้ใช้งานจริง รวมทั้งจัดทำรายงานที่จะเป็นประโยชน์ต่อการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ของกรมพินิจและคุ้มครองเด็กและเยาวชน

3. การออกแบบปัจจัยนำเข้าเพื่อใช้ในการสร้างตัวแบบการพยากรณ์การกระทำความผิดและกระทำความผิดซ้ำของเด็กและเยาวชน

จากการรวบรวมและวิเคราะห์ข้อมูลในการกระทำความผิดของเด็กและเยาวชนได้มีการนำข้อมูลจากระบบการประเมินความเสี่ยง (Risk: R) ของกรมพินิจและคุ้มครองเด็กและเยาวชนมาใช้เป็นปัจจัยนำเข้าเพื่อการเรียนรู้ แล้วสร้างตัวแบบการพยากรณ์การกระทำความผิดซ้ำของเด็กและเยาวชน ซึ่งประกอบด้วยปัจจัยนำเข้าทั้งสิ้น 46 ปัจจัย และผลลัพธ์จากการพยากรณ์ที่ได้จากการนำปัจจัยนำเข้าเหล่านี้ ประกอบด้วย 2 ผลลัพธ์ คือ มีการกระทำความผิดซ้ำ

และไม่มีการกระทำผิดซ้ำโดยเมื่อพบว่าเด็กและเยาวชนมีการกระทำผิดซ้ำ ระบบจะสามารถพยากรณ์ต่อไปได้ว่าการกระทำผิดซ้ำนั้นจะเป็นการกระทำตามฐานความผิดหลักข้อใด ดังแสดงในภาพที่ 5



ภาพที่ 5 แนวคิดในการสร้างต้นแบบเพื่อการพยากรณ์การกระทำผิดของเด็กและเยาวชน

จากภาพที่ 5 การกระทำผิดซ้ำและตามฐานความผิดหลัก เป็นผลลัพธ์มาจากปัจจัยนำเข้า ซึ่งผลลัพธ์ในปัจจุบันจะมีความสัมพันธ์กับผลลัพธ์ ณ ช่วงเวลาก่อนหน้า โดยลักษณะดังกล่าวเรียกว่า ความสัมพันธ์แบบเวียนบังเกิด (Recurrent relation) การเรียนรู้วิธีนี้มีข้อดี คือ มีการเรียนรู้ข้อผิดพลาดในอดีต แล้วนำมาปรับปรุงผลลัพธ์ในปัจจุบันให้ดียิ่งขึ้น อีกทั้งวิธีการนี้สามารถปรับลดหรือเพิ่มปัจจัยใหม่เข้าไปในระบบได้ง่าย

ผลการวิจัย

1. การทดสอบประสิทธิภาพของต้นแบบการพยากรณ์

การวัดประสิทธิภาพของต้นแบบการพยากรณ์ในการทำนายการกระทำผิดซ้ำใช้ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถูกต้อง (Accuracy) ซึ่งค่าที่ได้จะนำมาพิจารณาในตาราง Confusion Matrix ดังตารางที่ 1

ตารางที่ 1 ConfusionMatrix

		ค่าพยากรณ์	
		Class A (y=0)	Class B (y=1)
ค่าจริง	Class A (y=0)	True Positive (TP)	False Positive (FP)
	Class B (y=1)	False Negative (FN)	True Negative (TN)

จากตารางที่ 1 สมมุติให้ Class A เป็นกลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำ Class B เป็นกลุ่มเด็กและเยาวชนที่กระทำความผิดครั้งแรก และค่าต่าง ๆ ในตารางมีความหมาย ดังนี้

True Positive (TP) คือ จำนวนครั้งที่พยากรณ์ว่า เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำ และจากข้อมูลพบว่า เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำจริง

True Negative (TN) คือ จำนวนครั้งที่พยากรณ์ว่า ไม่เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำ และจากข้อมูลพบว่า ไม่เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำจริง

False Positive (FP) คือ จำนวนครั้งที่พยากรณ์ว่า เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำ และจากข้อมูลพบว่า ไม่เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำจริง

False Negative (FN) คือ จำนวนครั้งที่พยากรณ์ว่า ไม่เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำ และจากข้อมูลพบว่า เป็น กลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำจริง

ต้นแบบการพยากรณ์จะถูกนำไปประเมินประสิทธิภาพ เพื่อประเมินความถูกต้องและความแม่นยำของแบบจำลอง สามารถประเมินได้ 4 วิธี คือ

(1) ค่าความถูกต้อง (Accuracy) เป็นการวัดความถูกต้องของต้นแบบพยากรณ์โดยค่าที่ได้จะแสดงให้เห็นว่าต้นแบบการพยากรณ์ที่ใช้ในการพยากรณ์สามารถพยากรณ์ได้ว่าเป็นกลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำ และกลุ่มเด็กและเยาวชนที่กระทำความผิดได้ถูกต้องเพียงใด โดยคำนวณได้จาก

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

(2) ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำของต้นแบบการพยากรณ์ โดยค่าที่ได้จะแสดงให้เห็นว่าต้นแบบที่ใช้ในการพยากรณ์ สามารถพยากรณ์ว่าเป็นกลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำได้ถูกต้องเพียงใด โดยคำนวณได้จาก

$$Precision = \frac{TP}{TP + FP} \times 100$$

(3) ค่าความระลึก (Recall) เป็นการวัดความครอบคลุม โดยค่าที่ได้จะแสดงให้เห็นว่าตัวแบบที่ใช้ในการพยากรณ์สามารถพยากรณ์ว่าเป็นกลุ่มเด็กและเยาวชนที่กระทำความผิดซ้ำออกมาได้ครอบคลุมคำตอบที่มีทั้งหมดได้มากน้อยเพียงใด โดยคำนวณได้จาก

$$Recall = \frac{TP}{TP + FN} \times 100$$

2. ผลการพยากรณ์การกระทำความผิดของเด็กและเยาวชนโดยใช้ต้นแบบการพยากรณ์ด้วยอัลกอริธึม XGBoost

ผู้วิจัยทดสอบต้นแบบการพยากรณ์การกระทำความผิดของเด็กและเยาวชน ด้วยอัลกอริธึม XGBoost โดยมีการสร้างสถานการณ์จำลอง (Scenario) เพื่อวิเคราะห์ปัจจัยที่เป็นสาเหตุการกระทำความผิดของเด็กและเยาวชน สถานการณ์จำลองที่ได้กำหนดสำหรับการทดลอง ประกอบด้วย 4 สถานการณ์ ดังนี้

สถานการณ์ที่ 1 พยากรณ์การกระทำความผิดของเด็กและเยาวชน โดยใช้ปัจจัยที่ผ่านการคัดเลือกจากอัลกอริทึม XGBoost ประกอบด้วยปัจจัยดังต่อไปนี้ คือ Rsb6, Rdb11, Rd4, Rc2, Rc3 และ Rc4

สถานการณ์ที่ 2 พยากรณ์การกระทำความผิดของเด็กและเยาวชน โดยพิจารณาจากปัจจัยที่เป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชนที่มีความสัมพันธ์กันมากที่สุดในแต่ละด้าน (ค่าสัมประสิทธิ์สหสัมพันธ์ ตั้งแต่ 0.3 ขึ้นไป) ประกอบด้วยปัจจัยดังต่อไปนี้ คือ Rf1, Rf2, Rf3, Rf5, Rsb3, Rsb4, Rsb9, Rsb10, Rc2, Rc3, Rdb4, Rdb7, Rdb12, Rdb13 และ Rdb15

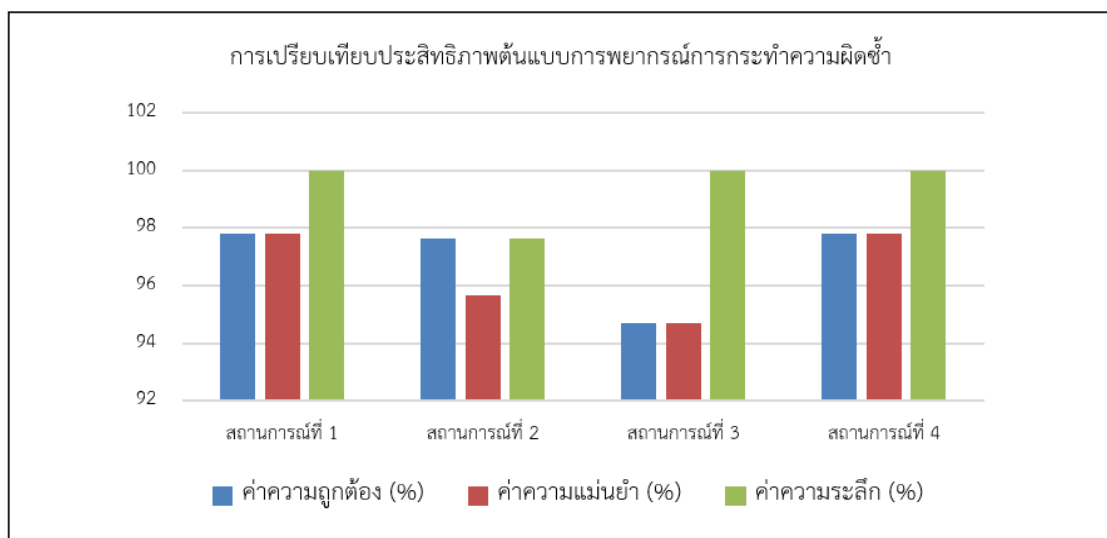
สถานการณ์ที่ 3 พยากรณ์การกระทำความผิดของเด็กและเยาวชน โดยใช้ปัจจัยหลัก ทั้ง 7 ด้าน

สถานการณ์ที่ 4 พยากรณ์การกระทำความผิดของเด็กและเยาวชน โดยพิจารณาเฉพาะปัจจัยที่มีคุณลักษณะสอดคล้องกับการกระทำความผิดของเด็กและเยาวชน ประกอบด้วยคุณลักษณะต่อไปนี้ คือ Rf1, Rf2, Rf5, Rce1, Rsb3, Rsb4, Rsb9, Rsb10, Rsb11, Rp1, Rp2, Rd1, Rd2, Rc1, Rc2 และ Rc4

การประเมินประสิทธิภาพของต้นแบบการพยากรณ์ใช้วิธีการทดสอบไขว้ แบ่งข้อมูลออกเป็น 5 ส่วน (5-folds cross validation) ผลการพยากรณ์พิจารณาจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision และค่าความระลึก (Recall) ผลลัพธ์จากการทดสอบต้นแบบการพยากรณ์เป็น ดังตารางที่ 2 และภาพที่ 6

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพของต้นแบบการพยากรณ์การกระทำความผิดซ้ำ ตามสถานการณ์จำลอง

สถานการณ์	ค่าความถูกต้อง (%)	ค่าความแม่นยำ (%)	ค่าความระลึก (%)
สถานการณ์ที่ 1	97.80	97.80	100.00
สถานการณ์ที่ 2	97.65	95.65	97.65
สถานการณ์ที่ 3	94.70	94.70	100.00
สถานการณ์ที่ 4	97.80	97.80	100.00



ภาพที่ 6 การเปรียบเทียบประสิทธิภาพต้นแบบการพยากรณ์การกระทำความผิดซ้ำ

จากตารางที่ 2 และภาพที่ 6 นำเสนอผลการประเมินประสิทธิภาพต้นแบบการพยากรณ์การกระทำความผิดซ้ำตามสถานการณ์จำลอง พบว่า สถานการณ์ที่ 1 และสถานการณ์ที่ 4 ให้ผลการประเมินประสิทธิภาพที่ดีที่สุดในทุกด้าน รองลงมา คือ สถานการณ์ที่ 2 และสถานการณ์ที่ 3 ตามลำดับ โดยโมเดลต้นแบบของสถานการณ์ที่ 1 และ 4 สามารถพยากรณ์การกระทำความผิดซ้ำหรือไม่ซ้ำ ได้อย่างถูกต้องร้อยละ 97.8 นอกจากนี้โมเดลต้นแบบสามารถพยากรณ์ได้แม่นยำ คิดเป็นร้อยละ 97.8 และค่าความระลึกราคาตอบทั้งหมดคิดเป็นร้อยละ 100

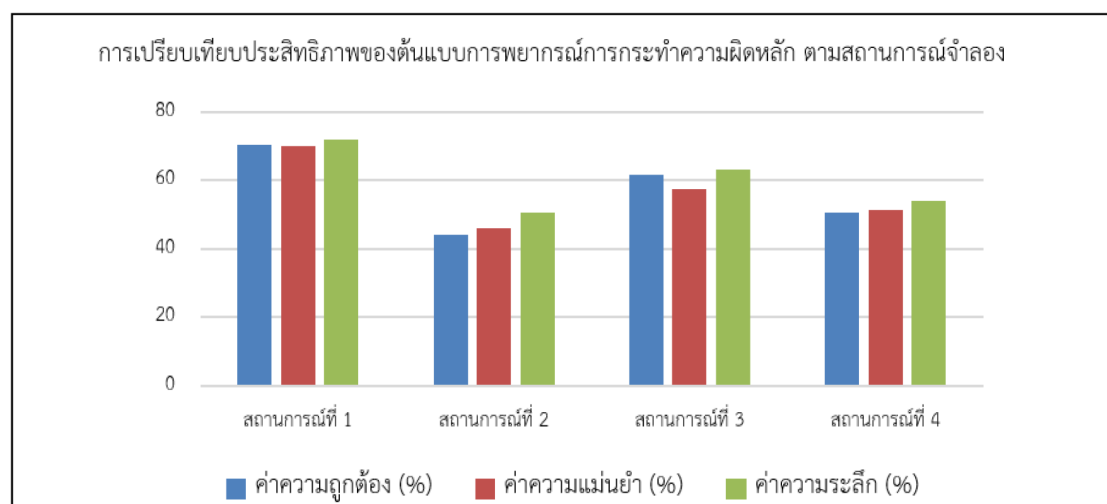
เมื่อทราบผลการพยากรณ์การกระทำความผิดตามสถานการณ์จำลองที่กำหนดแล้ว ระบบยังสามารถพยากรณ์ฐานความผิดหลักที่เด็กและเยาวชนถ้าได้กระทำความผิดแล้วจะกระทำความผิดซ้ำในฐานความผิดใด ฐานความผิดหลัก ประกอบด้วย

- P1: ฐานความผิดเกี่ยวกับยาเสพติดให้โทษ
- P2: ฐานความผิดเกี่ยวกับเพศ
- P3: ฐานความผิดเกี่ยวกับอาวุธและวัตถุระเบิด
- P4: ฐานความผิดเกี่ยวกับชีวิตและร่างกาย
- P5: ฐานความผิดเกี่ยวกับทรัพย์สิน
- P6: ฐานความผิดเกี่ยวกับความสงบสุข เสรีภาพ ชื่อเสียงและการปกครอง
- P7: ฐานความผิดอื่น ๆ

ผู้วิจัยทำการประเมินประสิทธิภาพต้นแบบการพยากรณ์ฐานความผิดหลัก ตามสถานการณ์จำลอง 4 สถานการณ์ที่ได้กำหนดไว้แล้ว ผลลัพธ์จากการทดสอบเป็น ดังตารางที่ 3 และภาพที่ 7

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพต้นแบบการพยากรณ์ฐานความผิดหลัก

สถานการณ์	ค่าความถูกต้อง (%)	ค่าความแม่นยำ (%)	ค่าความระลึก (%)
สถานการณ์ที่ 1	70.49	70.13	71.73
สถานการณ์ที่ 2	44.11	46.15	50.63
สถานการณ์ที่ 3	61.44	57.51	62.98
สถานการณ์ที่ 4	50.55	51.42	53.85



ภาพที่ 7 การเปรียบเทียบประสิทธิภาพของต้นแบบการพยากรณ์การกระทำความผิดตามสถานการณ์จำลอง

จากตารางที่ 3 และภาพที่ 7 นำเสนอผลการประเมินประสิทธิภาพต้นแบบการพยากรณ์การกระทำความผิดหลัก ตามสถานการณ์จำลอง พบว่า สถานการณ์ที่ 1 ให้ผลการประเมินประสิทธิภาพดีที่สุดในทุกด้าน รองลงมาคือ สถานการณ์ที่ 3 สถานการณ์ที่ 4 และสถานการณ์ที่ 2 ตามลำดับ

สรุปผลวิจัย

งานวิจัยนี้สร้างต้นแบบการพยากรณ์การกระทำความผิดซ้ำของเด็กและเยาวชน ด้วยอัลกอริธึม XGBoost โดยมีการกำหนดสถานการณ์จำลองในการพยากรณ์ที่ใช้จำนวนปัจจัยที่เป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชนแตกต่างกัน พบว่า การพยากรณ์โดยใช้ปัจจัยที่คัดเลือกโดยอัลกอริธึม XGBoost และเป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชนจำนวน 6 คุณลักษณะ และการพยากรณ์โดยใช้ปัจจัยที่เป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชน จำนวน 16 คุณลักษณะ มีผลการประเมินประสิทธิภาพของระบบด้วยค่าความถูกต้อง ค่าความแม่นยำ และค่าความระลึกลับ อยู่ในระดับสูงเช่นเดียวกัน กล่าวคือ ผลการประเมินด้านความถูกต้อง คิดเป็นร้อยละ 97.80 ค่าความแม่นยำ คิดเป็นร้อยละ 97.80 และค่าความระลึกลับ คิดเป็นร้อยละ 97.80 ดังนั้นในการพยากรณ์การกระทำความผิดซ้ำของเด็กและเยาวชน จึงสามารถพิจารณาโดยใช้จำนวนปัจจัยที่เป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชน จำนวน 6 และ 16 คุณลักษณะ แทนการพยากรณ์โดยใช้จำนวนปัจจัยที่เป็นสาเหตุของการกระทำความผิดของเด็กและเยาวชนทั้งหมด ก็สามารถให้ผลการพยากรณ์ที่มีประสิทธิภาพ เนื่องจากอัลกอริธึม XGBoost เป็นการเรียนรู้ที่นำความฉลาดเคลื่อนที่ได้จากการเรียนรู้ก่อนหน้ามาใช้ในการเรียนรู้ปัจจุบัน จึงทำให้มีความแม่นยำของการพยากรณ์มากขึ้นเรื่อย ๆ สอดคล้องกับงานวิจัยของวิไลลักษณ์ ตรีพิช (2562) ที่นำอัลกอริธึม XGBoost มาใช้ในการพยากรณ์จำนวนผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์ เช่นเดียวกับ Chen & Guestrin (2016) ที่นำอัลกอริธึม XGBoost ไปใช้ในการพยากรณ์ข้อมูลจาก KDD cup แล้วพบว่าผลการทำนายมีความถูกต้องอยู่ในระดับสูง

ข้อเสนอแนะ

มีการประยุกต์ใช้อัลกอริธึมประเภทอื่น ๆ และขยายผลการวิจัยสู่พื้นที่ต่าง ๆ เพื่อวิจัยคุณลักษณะ หรือปัจจัยต่าง ๆ ในแต่ละพื้นที่ ที่มีผลต่อการกระทำความผิดซ้ำของเด็กและเยาวชน โดยใช้ปัญญาประดิษฐ์ (Artificial Intelligence : AI) เข้ามาช่วยในการวิเคราะห์ และพยากรณ์การกระทำความผิดซ้ำตลอดจนเปรียบเทียบประสิทธิภาพของการพยากรณ์ในแต่ละอัลกอริธึม

เอกสารอ้างอิง

- โกศล วงศ์สวรรค์ และคณะ. (2537). ปัญหาสังคม. กรุงเทพฯ: รวมสารัน.
- กรมพินิจและคุ้มครองเด็กและเยาวชน (2562). สถิติคดีย้อนหลังระหว่างปี พ.ศ. 2553 – 2558. สืบค้น 6 กันยายน 2562, จาก <http://www.djop.go.th/Djop/main.php?page=NewsAll&CatID=63>.
- ชื่นจิตร วาณิชยสุวรรณ. (2554). ปัจจัยที่มีอิทธิพลต่อการกระทำความผิดของเด็กและเยาวชนในสถานพินิจและคุ้มครองเด็กและเยาวชนจังหวัดสงขลา. (วิทยานิพนธ์มหาบัณฑิต). คณะรัฐศาสตรมหาวิทยาลัยสงขลานครินทร์. สงขลา.
- ชินพัฒน์ แก้วชินพร. (2553). การจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจและการจัดกลุ่ม. วิทยานิพนธ์มหาบัณฑิต). สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง). กรุงเทพฯ.

- ชาญคณิต กฤตยา สุริยะมณี, และ อุนิษา เลิศโตมรสกุล (2553). ปัจจัยที่มีผลต่อการกระทำผิดซ้ำคดียาเสพติด: กรณีศึกษา ผู้กระทำผิดร้ายแรงและมีโทษสูงและต้องดูแลเป็นพิเศษ. กรุงเทพมหานคร: สำนักงานคณะกรรมการป้องกันและปราบปรามยาเสพติด กระทรวงยุติธรรม.
- นางเยาว์ ในอรุณ. (2557). การประยุกต์ใช้เทคนิคเหมืองข้อมูลในการพยากรณ์โรคเบาหวาน: กรณีศึกษาโรงพยาบาลศูนย์สวรรค์ประชากรวิทยานิพนธ์ ดุษฎีบัณฑิต. มหาวิทยาลัยนเรศวร. กรุงเทพมหานคร.
- นวลจันทร์ ทิศนชัยกุล. (2548). อาชญากรรม การป้องกัน: การควบคุม. นนทบุรี: พรทิพย์การพิมพ์.
- ปรารถนา อุดพัฒนา. (2560). แผนที่ความรุนแรงของอาชญากรรมในประเทศไทยจากการศึกษาด้วยวิธีเรียงลำดับความคล้ายคลึงกัน. การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษา ครั้งที่ 2, 20 มกราคม 2560 ณ มหาวิทยาลัยราชภัฏมหาสารคาม, (หน้า 408 – 412)
- วิไลลักษณ์ ตรีพิช. (2562). การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำเหมืองข้อมูล. วารสารวิจัยและพัฒนา วไลยอลงกรณ์ ในพระบรมราชูปถัมภ์, 14(1).
- สุชา จันทรเฒ. (2541). จิตวิทยาเด็กเกร. กรุงเทพฯ: ไทยวัฒนาพานิช.
- โสภา ซิลมณีน. (2536). บุคลิกภาพและพัฒนาการ: แนวโน้มสู่พฤติกรรมปกติ และการมีพฤติกรรมเบี่ยงเบนของเด็กและเยาวชน. กรุงเทพฯ: คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล.
- อนันต์ ประวิเศษ. (2547). ความคิดเห็นเกี่ยวกับโครงการบำบัด แก๊ซ ฟันฟู ผู้กระทำผิดของสถานพินิจและคุ้มครองเด็กและเยาวชน จังหวัดนครราชสีมา. (วิทยานิพนธ์รัฐประศาสนศาสตรมหาบัณฑิต สาขาวิชานโยบายสาธารณะ). มหาวิทยาลัยมหาสารคาม. มหาสารคาม.
- Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceeding of the 22nd ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, August, 2516, San Francisco California USA.
- Han, J., Kamber, M. & Pei, J. (2011). *Data Mining: Concepts and Techniques*. (3th ed.). Morgan Kaufmann Publisher.
- Zhang, D. & TSAI, J. P. (2003). Machine Learning and Software Engineering, Software Quality Journal, 11, 87–119.

คณะผู้เขียน

ผู้ช่วยศาสตราจารย์ ดร. ศิริลักษณ์ หล่อพันธุ์มณี

หลักสูตรวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์และเทคโนโลยี
295 มหาวิทยาลัยสวนดุสิต ถนนนครราชสีมา เขตดุสิต กรุงเทพมหานคร 10300
e-mail: siriluck_lor@dusit.ac.th

ผู้ช่วยศาสตราจารย์ ดร.ศิริพร ฉิมพลี

หลักสูตรวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์และเทคโนโลยี
295 มหาวิทยาลัยสวนดุสิต ถนนนครราชสีมา เขตดุสิต กรุงเทพมหานคร 10300
e-mail: siriporn_chi@dusit.ac.th

ผู้ช่วยศาสตราจารย์ ดร.วิชชา ฉิมพลี

หลักสูตรวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์และเทคโนโลยี
295 มหาวิทยาลัยสวนดุสิต ถนนนครราชสีมา เขตดุสิต กรุงเทพมหานคร 10300
e-mail: witcha_chi@dusit.ac.th

ผู้ช่วยศาสตราจารย์ วัชรกรณ์ เนตรหาญ

หลักสูตรวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์และเทคโนโลยี
295 มหาวิทยาลัยสวนดุสิต ถนนนครราชสีมา เขตดุสิต กรุงเทพมหานคร 10300
e-mail: eeeecon@yahoo.com

อาจารย์ณรงค์ฤทธิ์ ภิรมย์นง

หลักสูตรวิทยาการข้อมูลและการวิเคราะห์ คณะวิทยาศาสตร์และเทคโนโลยี
295 มหาวิทยาลัยสวนดุสิต ถนนนครราชสีมา เขตดุสิต กรุงเทพมหานคร 10300
e-mail: narongrit.piromnok@gmail.com