

การวิเคราะห์ความเสี่ยงพื้นที่น้ำท่วมโดยใช้แบบจำลองต้นไม้การตัดสินใจ บริเวณลุ่มน้ำทะเลสาบสงขลา

Flood Risk Analysis based on a Decision Tree Model at Songkhla Lake Basin

ธิดาภัทร อนุชาญ¹, นิตติ เอี่ยมชื่น²

Thidapath Anucharn¹, Niti lamchuen²

Received: 23 January 2017 ; Accepted: 5 May 2017

บทคัดย่อ

ลุ่มน้ำทะเลสาบสงขลา เป็นพื้นที่ที่เกิดน้ำท่วมและสร้างความเสียหายให้กับชีวิตและทรัพย์สิน ในการศึกษาครั้งนี้ได้ทำการวิเคราะห์หาปัจจัยที่มีอิทธิพลต่อการเกิดน้ำท่วม โดยการนำปัจจัยที่มีอิทธิพลต่อการเกิดน้ำท่วมมาพิจารณาพร้อมกับพื้นที่น้ำท่วมที่เคยเกิดขึ้นในอดีต ตั้งแต่ปี พ.ศ. 2550-2557 โดยใช้แบบจำลองต้นไม้การตัดสินใจ จากการศึกษาพบว่าปัจจัยที่เกี่ยวข้อง 4 ปัจจัย ได้แก่ 1) การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน 2) การระบายน้ำของพื้นผิวดิน 3) ความสูงจากระดับทะเลปานกลาง และ 4) ความหนาแน่นของทางน้ำ จากนั้นทำการวัดประสิทธิภาพของแบบจำลองในการจำแนกผลที่ได้สร้างขึ้นมานั้น พบว่าค่าการจับกลุ่มความถูกต้อง, ค่าการจับกลุ่มความไม่ถูกต้อง, ค่าความแม่นยำของแบบจำลอง, ค่าความระลึกของแบบจำลอง, ค่าความเหวี่ยงของแบบจำลอง และค่าพื้นที่โค้งจาก ROC คือ 95.21%, 4.79%, 0.896, 0.813, 0.853 และ 0.895 ตามลำดับ

คำสำคัญ: น้ำท่วม ลุ่มน้ำทะเลสาบสงขลา ต้นไม้การตัดสินใจ แผนที่ความอ่อนไหว

Abstract

The Songkhla lake basin is an area where flooding causes damage to people and properties. In this study, the factors affecting flooding were analyzed. All factors were considered in areas that had experienced past flooding between 2007 and 2014 using a decision tree model. The study provided evidence that the factors influencing flooding comprised 1) land use and land cover, 2) surface drainage, 3) altitude, and 4) stream network density. The evaluation efficiency of the model was addressed. The outcomes could be divided into 6 categories: correct classification instances, incorrect classification instances, precision, recall, F-measure and receiver operating characteristics (ROC) curve were 95.21%, 4.79%, 0.896, 0.813, 0.853 and 0.895 respectively.

Keywords: Flood, Songkhla lake basin, Decision tree, Susceptibility mapping

บทนำ

“ลุ่มน้ำทะเลสาบสงขลา” เป็นพื้นที่ป่าต้นน้ำที่สำคัญของทะเลสาบสงขลา อยู่ในเขตจังหวัดสงขลา พัทลุงและนครศรีธรรมราช (Figure 1) ส่วนใหญ่อยู่ในแนวเทือกเขา 2 เทือกเขาที่สำคัญคือ แนวเทือกเขาวรรทัด กับแนวเขาเตี้ยๆ ที่เป็นส่วนหนึ่งของเทือกเขาสันกาลาคีรี ซึ่งทั้งสองเทือกเขามีพื้นที่เป็นป่าดิบชื้น

พื้นที่ภูเขาส่วนใหญ่จัดเป็นเขตอนุรักษ์พันธุ์สัตว์ป่าหรืออุทยานแห่งชาติ ถัดจากพื้นที่ป่าต้นน้ำเป็นบริเวณชายเขาและที่ราบลุ่มซึ่งเป็นที่อยู่อาศัยของ “คนลุ่มน้ำ” และทำการเกษตรมีลำคลองหลายสายไหลผ่านลงสู่ทะเลสาบสงขลา บริเวณโดยรอบทะเลสาบสงขลา มีทั้งแหล่งชุมชนที่เป็นชุมชนชนบทและชุมชนเมือง พื้นที่พรุ พื้นที่การเกษตร พื้นที่อุตสาหกรรมและ

¹ อาจารย์, สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหาดใหญ่ 90110

² อาจารย์, สาขาวิชาภูมิสารสนเทศศาสตร์ คณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยพะเยา 56000

¹ Lecturer, Information Technology Department, Faculty of Science and Technology, Hatyai University, 90110 E-mail: thidapath.a@gmail.com

² Lecture, Department of Geographic Information Science, School of Information and Communication Technology, University of Phayao, 56000 E-mail: niti018@hotmail.com

ป่าชายเลน¹ จากลักษณะภูมิประเทศดังกล่าวส่งผลให้ที่ผ่านมามีบริเวณลุ่มน้ำทะเลสาบสงขลาเกิดน้ำท่วมซ้ำซาก สำหรับเหตุการณ์ที่สร้างความเสียหายให้แก่ทรัพย์สินเป็นอย่างมากในรอบปีต่างๆ ที่เคยเกิดขึ้นในบริเวณลุ่มน้ำทะเลสาบสงขลา โดยจังหวัดสงขลา ตั้งแต่ปี พ.ศ. 2555-2557 พบว่ามีประชาชนเสียชีวิต 18 ราย ได้รับความเดือดร้อนทั้งสิ้น 17,935 คน 182,356 ครัวเรือน รวมมูลค่าความเสียหาย 63,189,500 บาท² สำหรับจังหวัดพัทลุง ตั้งแต่ปี พ.ศ. 2539-2548 พบว่ามีประชาชน ได้รับความเดือดร้อนทั้งสิ้น 395,605 คน 118,328 ครัวเรือน มีพื้นที่การเกษตร เสียหาย 245,618 ไร่ รวมมูลค่าความเสียหาย 55,595,306 บาท³ ล่าสุดในปี พ.ศ. 2559 จังหวัดสงขลา น้ำท่วมขังในพื้นที่ 10 อำเภอ 63 ตำบล 352 หมู่บ้าน มีผู้เสียชีวิต 2 ราย ผู้บาดเจ็บ 1 ราย ประชาชนได้รับผลกระทบ 24,246 ครัวเรือน 81,319 คน อพยพประชาชน 232 ครัวเรือน 455 คน ถนนเสียหาย 369 สาย และจังหวัดพัทลุง น้ำท่วมขังในพื้นที่ 11 อำเภอ 62 ตำบล 572 หมู่บ้าน 45 ชุมชน มีผู้เสียชีวิต 2 ราย ประชาชนได้รับผลกระทบ 31,142 ครัวเรือน 85,393 คน อพยพประชาชน 27 ครัวเรือน⁴

จากการรวบรวมงานวิจัยที่เกี่ยวข้องของนาถนเรศ อากาศสุวรรณ, ประมาณ เทพสงเคราะห์ และวรุฒม์ นาที³, สุพิชฌาย์ ธนารุณ⁵, สุณัฐพงศ์ สูงสุมาลัย⁶, และ พิระพิทย์ พิฆังคผล และคณะ⁷ พบว่ามีหลายปัจจัยที่ นำมาใช้ในการวิเคราะห์หาพื้นที่น้ำท่วม จึงทำการคัดเลือกเพื่อกำหนดเป็นปัจจัยทางกายภาพที่มีอิทธิพลต่อการเกิดน้ำท่วมในพื้นที่ศึกษารวม 6 ปัจจัย ได้แก่ 1) ความสูงจากระดับน้ำทะเลปานกลาง 2) ความลาดชัน 3) การระบายน้ำของพื้นผิวดิน 4) ความหนาแน่นของทางน้ำ 5) ความหนาแน่นของเส้นทางถนน และ 6) การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน โดยมีรายละเอียดดังนี้

ความสูงจากระดับทะเลปานกลาง (Altitude: Mean sea level: MSL) ธรรมชาติของการไหลของน้ำจะไหลจากที่สูงลงสู่ที่ต่ำกว่า ตามลักษณะของพื้นที่ภาคใต้ตอนล่างมีแนวทิวเขาบรรทัดกั้นระหว่างฝั่งตะวันตกและฝั่งตะวันออก ทำให้มีความสูงลดหลั่นจากแนวภูเขาตอนกลางแผ่นดินไต้ระดับลงมาที่ราบฝั่งตะวันออกหรือฝั่งอ่าวไทย และได้รับความชุ่มชื้นจากอิทธิพลของทั้งจากลมมรสุมตะวันออกเฉียงเหนือในช่วงกลางเดือนตุลาคมถึงกลางเดือนกุมภาพันธ์ พัดผ่านอ่าวไทย ทำให้ช่วงดังกล่าวจะมีฝนตกชุกมาก ถ้าน้ำทะเลหนุนในช่วงนี้จะทำให้เกิดน้ำท่วมในพื้นที่ที่เป็นที่ต่ำได้ง่าย⁸ รายละเอียดดัง Figure 2

ความลาดชันของพื้นที่ (Slope) หมายถึง ลักษณะของพื้นที่ที่ลาดเอียงไปจากแนวราบ หรือสัดส่วนของระยะใน

แนวตั้งกับแนวนอน ความลาดชันของพื้นที่จะมีอิทธิพลต่อลักษณะสมบัติของดิน การไหลบ่าของน้ำและตะกอน พื้นที่ที่มีความลาดชันต่ำกว่า หรือเป็นที่ราบเมื่อเกิดฝนตกบนพื้นที่นั้นก็จะทำให้เกิดน้ำท่วมขังในพื้นที่นั้นเป็นระยะเวลานาน แต่ในทางตรงกันข้ามหากพื้นที่ที่มีความลาดชันสูง เมื่อเกิดฝนตกลงมาก็จะทำให้หน้าฝนไหลออกจากพื้นที่นั้นได้ภายในระยะเวลาที่ไม่นาน ดังนั้นพื้นที่ที่มีความลาดชันต่ำ ก็มีความเสี่ยงต่อการเกิดน้ำท่วมสูงกว่าพื้นที่ที่มีความลาดชันสูง รายละเอียดดัง Figure 3

การระบายน้ำของพื้นผิวดิน (Surface Drainage)

เนื้อดินที่มีผลต่อการเกิดอุทกภัย คือ ดินที่มีเนื้อละเอียดมาก โดยเฉพาะดินที่มีค่าอนุภาคของดินเหนียวสูงจะมีการระบายน้ำไม่ดี ทำให้เกิดการแช่ขังของน้ำ ในทางกลับกันดินที่มีอนุภาคของดินเหนียวต่ำจะมีการระบายน้ำได้ดี ซึ่งจะช่วยให้การระบายน้ำส่วนเกินออกจากพื้นที่ประสบน้ำท่วมเป็นไปอย่างรวดเร็วทำให้ไม่เกิดการแช่ขังของน้ำ โดยได้จัดกลุ่มความสามารถในการระบายน้ำของดินตามศักยภาพในการระบายน้ำอ้างอิงตามกรมพัฒนาที่ดิน⁹ ออกเป็น 8 ระดับ (Figure 4) ดังนี้ การระบายน้ำเร็วมาก, การระบายน้ำเร็ว, การระบายน้ำค่อนข้างเร็ว, การระบายน้ำดี, การระบายน้ำดีปานกลาง, การระบายน้ำค่อนข้างมาก, การระบายน้ำมากเกินไป และพื้นที่ไม่มีการสำรวจ (ความลาดชันเชิงซ้อน หรือ Slope complex จัดอยู่ในกลุ่มชุดดิน 62)

ความหนาแน่นของทางน้ำ (Stream Density)

ทางน้ำเกิดจากรวมตัวของน้ำที่ไหลรวมกัน ในพื้นที่มีลักษณะรูปแบบทางน้ำแบบขนนก (Dendritic Pattern) โดยพื้นที่ลุ่มน้ำใดที่มีทางน้ำหนาแน่น การระบายน้ำของพื้นที่ก็จะดี โอกาสที่จะเกิดน้ำท่วมก็จะน้อยตามไปด้วย การพิจารณาความหนาแน่นของทางน้ำ รายละเอียดดัง Figure 5

ความหนาแน่นของเส้นทางคมนาคม (Road Density) ลักษณะของถนนในพื้นที่ส่วนใหญ่เป็นลักษณะเส้นทางเพื่อเชื่อมต่อบริเวณระหว่างชุมชนใกล้เคียงและในลำดับขั้นของชุมชนที่สูงขึ้นไปแบบโครงข่าย (Road Network) ในบริเวณพื้นที่ที่มีเส้นทางคมนาคมมากก็จะเป็นอุปสรรคต่อการไหลของน้ำ ส่งผลต่อการระบายน้ำออก เพราะมีสิ่งกีดขวาง ขวางเส้นทางการไหลของน้ำ และทำให้การระบายน้ำส่วนเกินออกจากพื้นที่ที่เป็นไปอย่างล่าช้า โอกาสเกิดอุทกภัยก็จะเพิ่มสูงขึ้นตามไปด้วย รายละเอียดดัง Figure 6

การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (Land Use and Land Cover: LULC) ลักษณะของสิ่งปกคลุมดินหรือการใช้ที่ดินส่งผลต่อการเกิดน้ำท่วมในบริเวณพื้นที่ที่มีพืชพรรณปกคลุมดิน เช่น พื้นที่ไถมีป่าไม้ ไม้ยืนต้น และสวนผล

ไม่อยู่มาก พื้นที่นั้นจะมีโอกาสเกิดน้ำท่วมน้อย เนื่องจากพืชพรรณดังกล่าวสามารถช่วยดูดซับน้ำได้ดี ทำให้หน้าส่วนเกินมีปริมาณลดลงและยังช่วยชะลอการไหลบ่าของน้ำ ด้วยการใช้ประโยชน์ที่ดินที่มีผลต่อการเกิดน้ำท่วม จึงพิจารณาจากชนิดของสิ่งที่ปกคลุมดินในระดับที่ 1 โดยจัดลำดับออกเป็น 5 ระดับ

(Figure 7) อ้างอิงตามกรมพัฒนาที่ดิน¹⁰ ดังนี้ พื้นที่การใช้ประโยชน์อื่น ๆ (พื้นที่ลุ่ม, ไม้พุ่มหรือทุ่งหญ้าสลัปไม้พุ่ม, เหมือนแระ เหมือนเก่า บ่อขุด), พื้นที่อยู่อาศัย, แหล่งน้ำ, พื้นที่เกษตรกรรม และพื้นที่ป่าไม้



Figure 1 Province Political Boundary : Songkhla Lake Basin

วิธีการที่นำมาใช้ในการวิเคราะห์ปัจจัยดังกล่าวส่วนใหญ่จะเป็นการให้ค่าน้ำหนักคะแนนความเหมาะสมของปัจจัย (Weighting) และการให้ค่าน้ำหนักคะแนนระดับของปัจจัย (Rating)^{3,5,6,7} แต่สำหรับการศึกษาคั้งนี้ ได้นำวิธีโครงสร้างต้นไม้การตัดสินใจ (Decision Tree) มาประยุกต์ใช้ โดยโครงสร้างต้นไม้การตัดสินใจจะประกอบด้วย 1) ใบ (Leaf) เป็นส่วนของผลลัพธ์หรือกลุ่ม ที่ได้กำหนดไว้ 2) กิ่ง (Branch) เป็นตัวเชื่อมระหว่างโหนด ที่แสดงผลในการทดสอบ (ในที่นี้โหนดจะแสดงแอตทริบิวต์ข้อมูล (Attribute Data) 3) ราก (Root) เป็นจุดเริ่มต้นของเหตุการณ์¹¹ การสร้างจะเป็นลักษณะจากบนลงล่าง (Top-Down) คือ เริ่มจากการสร้างรากของต้นไม้ก่อน แล้วแตกแขนงไปตามเงื่อนไขหรือข้อมูลที่ได้อาคาดการณ์ไว้ว่าจะเกิดขึ้น สำหรับข้อดีของแบบจำลองนี้ คือ เป็นแบบจำลองที่เข้าใจง่าย นำไปสร้างกฎแสดงความสัมพันธ์ของข้อมูลได้ง่าย มีลักษณะเหมือนแผนภูมิองค์กร สามารถแปลความจากแบบจำลองได้เลย แบบจำลองที่สร้างได้ อาจจะไม่ใช้ทุกแอตทริบิวต์ในชุดข้อมูลฝึกสอน (Training Data) และสามารถวิเคราะห์ปัญหาที่ซับซ้อนได้ ข้อด้อย คือ แบบจำลองที่ได้ อาจมีขนาด

ใหญ่ซับซ้อน การประยุกต์ใช้งานจะเป็นลักษณะของการเรียนรู้แบบมีผู้สอน (Supervise Learning) คือ สามารถนิยามได้ว่าผลลัพธ์ จะเป็นอย่างไร เช่น ผ่าน/ไม่ผ่าน ,ใช่/ไม่ใช่ มีประโยชน์ในการสร้างโมเดลแบบการพยากรณ์ (Forecasting) สามารถใช้ได้หลากหลาย เช่น สถาบันการเงินนำมาใช้ในการอนุมัติเครดิตว่าบุคคลคนนี้จะเครดิตผ่าน/ไม่ผ่าน หรือการเก็บข้อมูลสภาพภูมิอากาศ ย้อนหลังเพื่อดูว่าฝนจะตกหรือไม่ ดังนั้นผู้วิจัยจึงได้นำวิธีการดังกล่าวมาใช้ในการพยากรณ์การเกิดน้ำท่วม ว่าพื้นที่ใดบ้างท่วม/ไม่ท่วม

สำหรับการวิเคราะห์ต้นไม้การตัดสินใจใช้ซอฟต์แวร์ Weka ซึ่งย่อมาจาก Waikato Environment for Knowledge Analysis¹² ซึ่งเป็นซอฟต์แวร์สำเร็จรูปประเภทซอฟต์แวร์โอเพ่นซอร์ส (Free Ware) ที่สามารถดาวน์โหลดได้จากเว็บไซต์ซึ่งอยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม Weka ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด

กรอบแนวคิดในการศึกษาคั้งนี้ ได้ทำการวิเคราะห์หาปัจจัยที่มีอิทธิพลต่อการเกิดน้ำท่วม และทำการสร้างแผนที่ความอ่อนไหวต่อการเกิดน้ำท่วม โดยใช้แบบจำลองต้นไม้การ

ตัดสินใจ ซึ่งระบุพื้นที่ที่มีแนวโน้มการเกิดน้ำท่วมบนพื้นฐานความรู้ของตำแหน่งที่เคยเกิดน้ำท่วมในอดีตกับปัจจัยทางกายภาพที่มีความสัมพันธ์ที่ส่งผลต่อการเกิดน้ำท่วม โดย

แผนที่ที่ได้จะนำมาใช้ประกอบการวางแผนบรรเทาปัญหาอุทกภัย การแจ้งเตือนภัย และการเตรียมรับสถานการณ์ช่วงฤดูฝนในบริเวณลุ่มน้ำทะเลสาบสงขลา



Figure 2 Allitude: Mean Sea Level



Figure 3 Slope

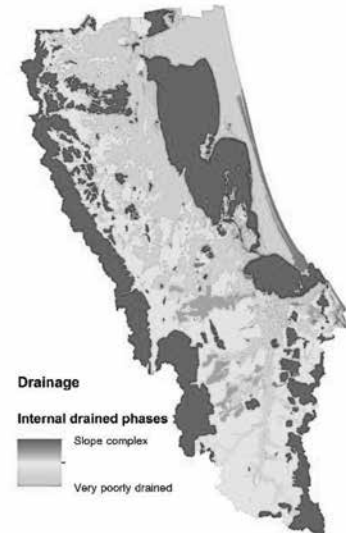


Figure 4 Surface Drainage

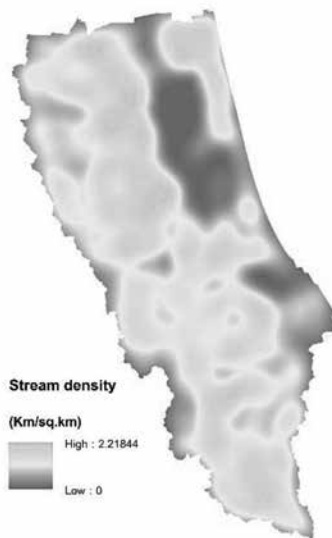


Figure 5 Stream Density



Figure 6 Road Density

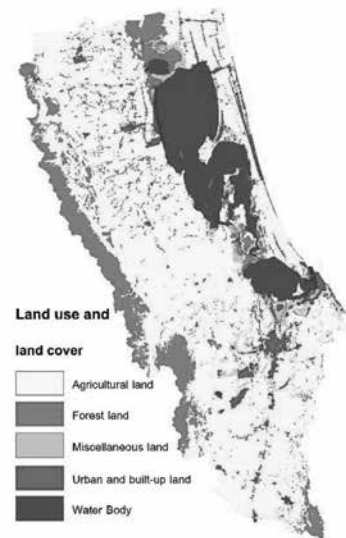


Figure 7 Land Use and Land Cover

วิธีการศึกษา

นำข้อมูลการเกิดน้ำท่วมตั้งแต่ปี พ.ศ. 2550 – 2557 เป็นระยะเวลา 8 ปี (Figure 8) จากระบบติดตามน้ำท่วมของสำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน)¹³ และปัจจัยทางกายภาพทั้ง 6 ปัจจัย (Table 1 และ 2) ไปวิเคราะห์หาความสัมพันธ์โดยใช้แบบจำลองต้นไม้การตัดสินใจ ซึ่งขั้นตอนการสร้างต้นไม้การตัดสินใจ มี 6 ขั้นตอนคือ 1) เริ่มจากรากที่ประกอบด้วยโหนดเพียงโหนดเดียว แสดงถึงชุดข้อมูลที่ใช้ฝึกสอน 2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้วให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามประเภทหรือกลุ่มของข้อมูลนั้น 3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่จะต้องวัดค่าสัดส่วนเกน (Gain Ratio) ของแต่ละแอตทริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอตทริบิวต์ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่างๆ ได้ดีที่สุด โดยแอตทริบิวต์ที่มีค่าสัดส่วนเกนมากที่สุด จะถูกเลือกให้เป็นตัวทดสอบหรือแอตทริบิวต์ที่ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้ 4) กิ่งของต้นไม้ถูกสร้างขึ้นจากค่าต่างๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่างๆ ที่สร้างขึ้น 5) ทำการวนซ้ำเพื่อหาแอตทริบิวต์ที่มีค่าสัดส่วนเกนมากที่สุดสำหรับข้อมูลที่ถูกแบ่งแยกออกมาใน

แต่ละกิ่ง เพื่อนำแอตทริบิวต์นี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่แอตทริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีกสำหรับโหนดในระดับต่อไป และ 6) ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อยๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งเป็นจริง คือ ข้อมูลที่ใช้ในการเรียนรู้ทุกตัวมีค่ากลุ่มเหมือนกันหมด หรือ ไม่มีแอตทริบิวต์เหลือในการแบ่ง หรือไม่มีข้อมูลที่ใช้ในการฝึกสอนเหลืออยู่¹⁴

สำหรับวิธีการแยกประเภทหรือกลุ่มข้อมูล ใช้วิธีการ J48 ซึ่งเป็นอัลกอริทึมที่ใช้หลักการของการใช้ทฤษฎีข่าวสาร (Information Theory) และค่าที่วัดได้จะนำมาใช้ตัดสินใจว่าจะใช้ตัวแปรใดในการทำนาย หรือแบ่งประเภทของข้อมูล โดยชุดตัวอย่าง (Sample) คือ ชุดของข้อมูลที่ใช้ฝึกสอน (Training Sample) เพื่อให้ข้อมูลเรียนรู้ และสร้างแบบจำลอง (Model Construction) และทดสอบโดยชุดข้อมูลสำหรับการทดสอบ (Testing Sample) เพื่อประเมินความถูกต้องของแบบจำลอง (Model Evaluation)¹⁵ ตัวแปรเป้าหมาย (Target Attribute) คือ ตัวแปรที่นำค่าไปใช้ในการทำนายผลในโครงสร้างต้นไม้ ในที่นี้คือ การทำนายผลน้ำท่วมหรือน้ำไม่ท่วม และแอตทริบิวต์คือ ตัวแปร (ปัจจัย) อื่นๆ ที่ใช้ในการสร้างโหนดในต้นไม้และไม่ใช้ตัวแปรเป้าหมาย

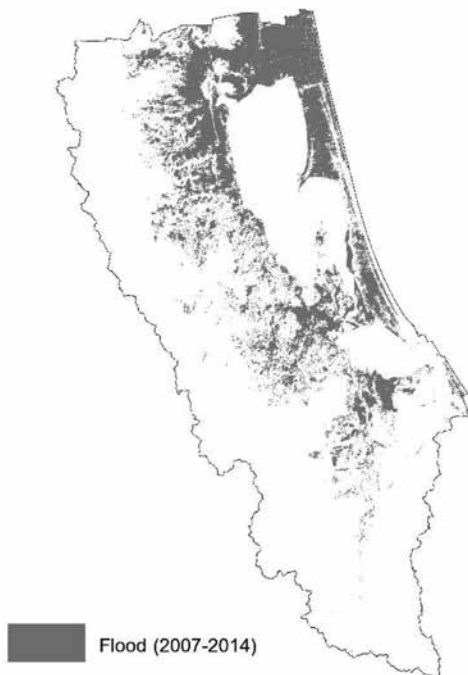


Figure 8 Flood Plan in the past (2007-2014)

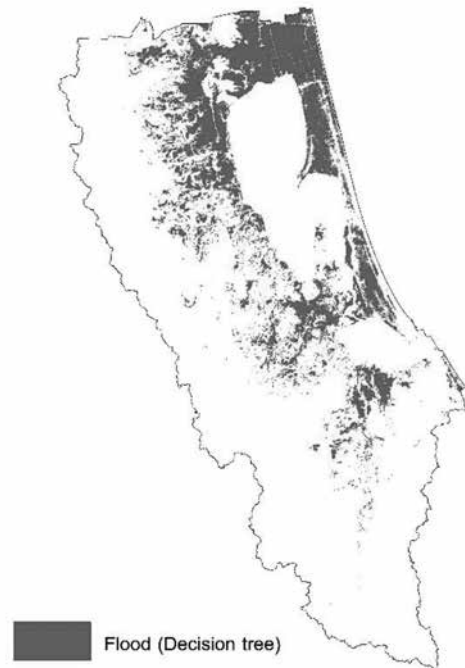


Figure 9 Flood Susceptibility Map

Table 1 Determine the data Types of each factor

แอตทริบิวต์	ความหมาย	ค่าที่เป็นไปได้ของแอตทริบิวต์
Flood	น้ำท่วม	0 = ไม่เกิดน้ำท่วม , 1 = เกิดน้ำท่วม
Altitude	ความสูงจากระดับทะเลปานกลาง	1 – 1,348 เมตร
Slope	ความลาดชัน	0 – 85 องศา
Surface Drainage	การระบายน้ำของพื้นผิวดิน	การระบายน้ำเลวมาก (1), การระบายน้ำเลว (2), การระบายน้ำค่อนข้างเลว (3), การระบายน้ำดี (4), การระบายน้ำดีปานกลาง (5), การระบายน้ำค่อนข้างมาก (6), การระบายน้ำมากเกินไป (7) และพื้นที่ไม่มีการสำรวจ (8)
Stream Density	ความหนาแน่นทางน้ำ	0 – 2.22 กิโลเมตร/ตารางกิโลเมตร
Road Density	ความหนาแน่นของเส้นทางคมนาคม	0 – 14.50 กิโลเมตร/ตารางกิโลเมตร
LULC	การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน	พื้นที่การใช้ประโยชน์อื่น ๆ (พื้นที่ลุ่ม, ไม้พุ่มหรือทุ่งหญ้าสลับไม้พุ่ม, เหมืองแร่ เหมืองเก่า บ่อขุด) (1), พื้นที่อยู่อาศัย (2), แหล่งน้ำ (3), พื้นที่เกษตรกรรม (4) และพื้นที่ป่าไม้ (5)

หมายเหตุ ค่าที่เป็นไปได้ของแอตทริบิวต์ ที่มี (1), (2),..., และ (8) ใช้เป็นตัวแทนในการนำเข้าสู่ข้อมูลประเภทข้อความ

Table 2 Example shows how to set data inputs for each factor

Flood	Altitude	Stream Density	Road Density	Slope	LULC	Surface Drainage
1	11.672	0.497	0.170	0.010	4	1
1	14.327	0.553	0.928	0.004	1	2
1	17.635	0.393	1.240	0.009	4	2
0	11.091	0.744	0.660	0.138	4	2
0	13.323	1.413	1.214	0.000	2	2
0	40.127	1.064	1.037	0.428	4	2

ในการสร้างโหนดในต้นไม้และไม่ใช้ตัวแปรเป้าหมาย (Target Attribute) J48 สามารถหลีกเลี่ยงการสร้างโครงสร้างต้นไม้ที่ใหญ่เกินไป เนื่องจากมีข้อมูลจำนวนมาก อย่างไรก็ตามขึ้นอยู่กับข้อกำหนดความลึกเมื่อมีการเติบโตของต้นไม้การตัดสินใจ นอกจากนี้ยังมีการรวมและกำจัดกิ่งที่มีผลกระทบต่อความถูกต้องที่น้อยออก มีการตัดทอนความผิดพลาดออกไป (Pruning Node) มีการสร้างกฎหลังการตัดทอนข้อมูลที่ผิดพลาดออก การเลือกแอตทริบิวต์ที่วัดการเลือกให้เหมาะสมสามารถใช้กับข้อมูลที่มีความต่อเนื่องที่เป็นตัวเลข และสามารถใช้กับชุดข้อมูลที่มีค่าผิดพลาด (Missing Attribute)¹⁶

วิธีการหาค่าสัดส่วนเกณฑ์นั้น จะต้องทำการหาค่า Entropy, Information Gain และ Split Information¹⁷ ซึ่งมีรายละเอียดดังนี้

การหาค่า Entropy เป็นสมการที่ใช้ในการหาค่าสารสนเทศของข้อมูล (Entropy Measure) รายละเอียดตั้งสมการที่ 1

$$Entropy(s) = \sum_{i=1}^c -P_i \log_2 P_i \quad (1)$$

โดย s คือ แอตทริบิวต์ หรือคุณลักษณะที่นำมาวัดค่า Entropy P_i คือ สัดส่วนของจำนวนสมาชิกในกลุ่ม i เทียบกับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

การหาค่า Information Gain เป็นสมการที่ใช้ในการหาค่าสารสนเทศก่อนนำไปใช้ในการหาค่ามาตรฐานอัตราส่วนเกณฑ์ (Gain Ratio) รายละเอียดตั้งสมการที่ 2

$$Gain(S, A) = Entropy(s) - \sum_{v \in \text{value}(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

โดย A คือ แอตทริบิวต์ A
 $|S_v|$ คือ จำนวนสมาชิกของแอตทริบิวต์ A ที่มีค่า V
 $|S|$ คือ จำนวนสมาชิกของกลุ่มตัวอย่าง

การหาค่า Split Information เป็นสมการที่ใช้ในการหาค่าสารสนเทศของการแบ่งแยก รายละเอียดดังสมการที่ 3

$$Split\ Information(S, A) = - \sum_{i=1}^n \frac{|S_i|}{S} \log_2 \frac{|S_i|}{|S|} \quad (3)$$

โดย S_i คือ จำนวนของจำนวนสมาชิกในกลุ่ม i การหาค่า Gain Ratio เพื่อลดความลำเอียงของข้อมูล รายละเอียดดังสมการที่ 4

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{Split\ Information(S, A)} \quad (4)$$

สำหรับการวัดประสิทธิภาพของแบบจำลองในการจำแนก ได้ใช้ 1) ค่าการจัดกลุ่มความถูกต้อง (Correctly Classification Instances) ในการยืนยันผลการทำนายข้อมูลว่ามีความถูกต้องร้อยละเท่าไรเมื่อเทียบกับข้อมูลทั้งหมด 2) ค่าการจัดกลุ่มความไม่ถูกต้อง (Incorrectly Classification Instances) โดยจะอธิบายเกี่ยวกับผลการทำนายข้อมูลที่ผิดพลาดเป็นร้อยละเทียบกับข้อมูลทั้งหมด 3) ค่าความแม่นยำของแบบจำลอง (Precision) 4) ค่าความระลึกของแบบจำลอง (Recall) 5) ค่าความเหวี่ยงของแบบจำลอง (F-Measure) 6) ผลการทดสอบความสามารถในการทำนายค่าคะแนนความเสี่ยง (Relative Operating Characteristic-ROC Curve) เป็นค่าที่เกิดจากการเปรียบเทียบระหว่างค่าที่ทำนายถูก (TP Rate) และค่าที่ทำนายผิด (FP Rate) ของแต่ละกลุ่มเป้าหมาย ซึ่งในการคำนวณค่าต่างๆ มีรายละเอียดดังนี้

$$Correctly\ Classified\ Instances = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

$$Incorrectly\ Classified\ Instances = \frac{FP + FN}{TP + FP + FN + TN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (9)$$

$$TP\ Rate = \frac{TP}{TP + FN} \quad (10)$$

$$FP\ Rate = \frac{FP}{FP + TN} \quad (11)$$

โดย
 TP คือ True Positive ค่ากลุ่มเป้าหมายคือ Yes และตัวแบบทำนายว่า Yes
 TN คือ True Negative ค่ากลุ่มเป้าหมายคือ No และตัวแบบทำนายว่า No
 FP คือ False Positive ค่ากลุ่มเป้าหมายคือ No และตัวแบบทำนายว่า Yes
 FN คือ False Negative ค่ากลุ่มเป้าหมายคือ Yes และตัวแบบทำนายว่า No

ผลการศึกษา

การวิเคราะห์ต้นไม้การตัดสินใจโดยใช้ซอฟต์แวร์ Weka ซึ่งใช้เทคนิควิธีการแบ่งประเภทข้อมูล (Classification) ด้วยวิธี J48 สามารถแสดงผลการวิเคราะห์ดัง Figure 10 ซึ่งพบว่าปัจจัยที่มีอิทธิพลต่อการเกิดน้ำท่วม มี 4 ปัจจัยเท่านั้น (จาก 6 ปัจจัย) คือ 1) การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน 2) การระบายน้ำของพื้นผิวดิน 3) ความสูงจากระดับทะเลปานกลาง และ 4) ความหนาแน่นของทางน้ำ โดยการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินที่ได้จัดลำดับไว้เป็น 5 ระดับ คือ พื้นที่การใช้ประโยชน์อื่นๆ (พื้นที่ลุ่ม, ไม้พุ่มหรือทุ่งหญ้าสลับไม้พุ่ม, เหมือนแระ เหมือนเก่า บ่อขุด), พื้นที่อยู่อาศัย, แหล่งน้ำ, พื้นที่เกษตรกรรม และพื้นที่ป่าไม้ มีอิทธิพลต่อการเกิดน้ำท่วมอย่างมาก ถัดไปคือการระบายน้ำของพื้นผิวดิน ซึ่งการระบายน้ำเร็วและการระบายน้ำช้า และพื้นที่ที่มีความลาดชันเชิงชัน มีอิทธิพลต่อการเกิดน้ำท่วมในลำดับถัดมา ในขณะที่ความสูงจากระดับทะเลปานกลางที่น้อยกว่าหรือเท่ากับ 40.339 เมตร และ ความหนาแน่นของทางน้ำที่มากกว่า 1.595 กิโลเมตรต่อตารางกิโลเมตร มีอิทธิพลต่อการเกิดน้ำท่วมเช่นเดียวกัน

ผลการวัดประสิทธิภาพของแบบจำลองในการจำแนกพื้นที่น้ำท่วมที่สร้างขึ้น มีรายละเอียดดังนี้ (1) Correctly Classification Instances ค่าที่ทำนายถูกต้องในการจำแนกกลุ่ม (2) Incorrectly Classification Instances ค่าที่ทำนายผิดพลาดในการจำแนกกลุ่ม ผลลัพธ์ที่ได้คือ 95.21% และ 4.79% ตามลำดับ (3) Precision ค่าของตัวแบบที่ทำนายได้ถูกต้อง ผลลัพธ์คือ 0.896 (4) Recall ค่าจากการทำนายตัวแบบที่ตรงกับความเป็นจริง ผลลัพธ์คือ 0.813 (5) F-Measure ค่าที่เกิดจากการเปรียบเทียบระหว่างค่าความแม่นยำ (Precision) และค่า Recall ของกลุ่มเป้าหมาย ผลลัพธ์คือ 0.853 และ (6) ROC Curve ค่าที่เกิดจากการเปรียบเทียบระหว่างค่า TP rate (ค่าที่

ทำนายถูก) และค่า FP rate (ค่าที่ทำนายผิด) ของแต่ละกลุ่มเป้าหมาย ผลลัพธ์คือ 0.895 (ค่าเข้าใกล้ 1 หมายความว่ามีความน่าเชื่อถือในการพยากรณ์¹⁸⁾ หรืออยู่ในเกณฑ์ระดับดี¹⁹ หลังจากพิจารณาเห็นว่าแบบจำลองมีความน่าเชื่อถือ จึงนำ

แบบจำลองต้นไม้มัดการตัดสินใจมาทำการวิเคราะห์ข้อมูลเชิงพื้นที่ เพื่อแสดงว่าบริเวณส่วนไหนมีความน่าจะเป็นในการเกิดน้ำท่วมหรือน้ำไม่ท่วม ซึ่งได้แสดงในแผนที่ความอ่อนไหวต่อการเกิดน้ำท่วมดัง Figure 9

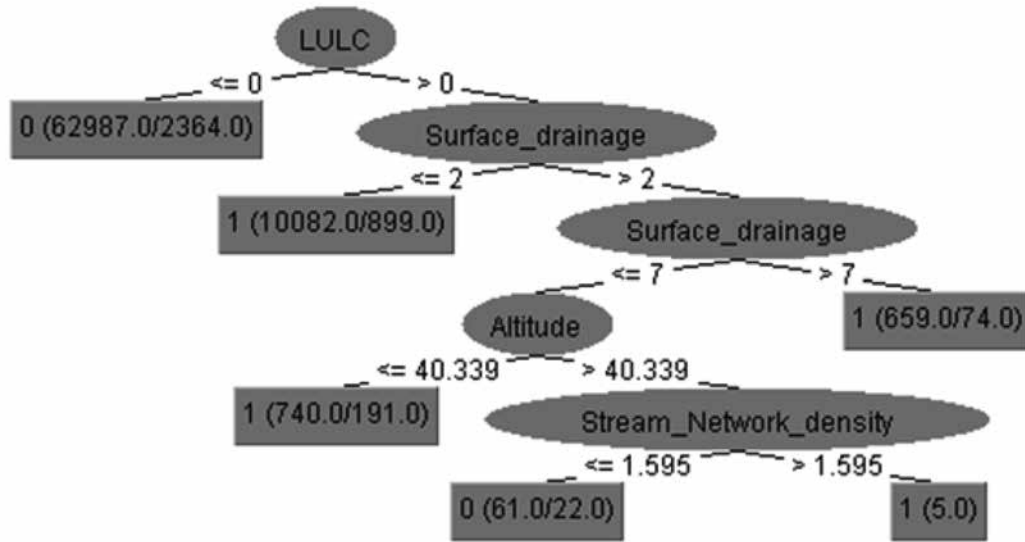


Figure 10 A Decision Tree Model Generated by: J48

วิจารณ์และสรุปผลการศึกษา

ผลการศึกษารวบรวมความเสี่ยงพื้นที่น้ำท่วม บริเวณลุ่มน้ำทะเลสาบสงขลาในภาคใต้ของประเทศไทย ได้นำปัจจัยที่เกี่ยวข้องนำมาวิเคราะห์ รวม 6 ปัจจัยจากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้อง ได้แก่ 1) การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน 2) การระบายน้ำของพื้นผิวดิน 3) ความสูงจากระดับทะเลปานกลาง และ 4) ความหนาแน่นของทางน้ำ 5) ความหนาแน่นของเส้นทางคมนาคม 6) ความลาดเทของพื้นที่ ภายหลังใช้แบบจำลองต้นไม้การตัดสินใจพบว่า มี 4 ปัจจัยที่สัมพันธ์กับความเสี่ยงการเกิดน้ำท่วม โดยปัจจัยที่ถูกตัดออก 2 ปัจจัย ได้แก่ ความหนาแน่นของเส้นทางคมนาคมและความลาดเทของพื้นที่ เมื่อทดสอบด้วยวิธีการทางสถิติคือการวิเคราะห์การถดถอยโลจิสติกแบบทวิ (Binary Logistic Regression Analysis: BLRA) พบว่า ค่าที่ได้จากการทดสอบมีค่าใกล้เคียงกัน กล่าวคือ เมื่อนำปัจจัยทั้ง 6 ปัจจัยมาวิเคราะห์ พบว่า ในรอบที่ 4 (Step) ของการทดสอบ คงเหลือ 4 ปัจจัยเช่นเดียวกับที่ได้จากแบบจำลองต้นไม้การตัดสินใจ

แบบจำลองต้นไม้การตัดสินใจช่วยลดตัวแปรที่ไม่มีอิทธิพลต่อการเกิดน้ำท่วมโดยมีค่าสถิติอ้างอิง และสามารถเข้ากับข้อมูลที่มีลักษณะเป็นมาตรการวัดแบบนามบัญญัติ (Nominal Data) ได้ และเมื่อทำการทวนสอบ (Validation) กับข้อมูล

น้ำท่วมจริงที่เกิดขึ้น (Figure 8,9) พบว่ามีค่าที่ทำนายถูกต้องสูงถึงร้อยละ 95 และมีค่าสถิติแคปป่า(Kappa Statistic) ถึง 82.41% ที่ได้จากการทดสอบด้วยซอฟต์แวร์ Weka นอกจากนี้ยังทำให้ทราบค่าที่ชัดเจนในการกำหนดพื้นที่เฉพาะ (Hot Spot) ที่มีความเสี่ยงจากข้อมูลที่ได้จากแบบจำลองต้นไม้การตัดสินใจในการนำไปสู่การวางแผนเฝ้าระวัง เช่น พื้นที่มีการระบายน้ำเลวถึงระบายน้ำมากเกินไป และความสูงที่น้อยกว่า 40.339 เมตร และมีความหนาแน่นของลำน้ำต่อพื้นที่มากกว่า 1.595 กิโลเมตรต่อตารางกิโลเมตร (ความยาวลำน้ำต่อพื้นที่) เป็นพื้นที่เสี่ยงเกิดน้ำท่วมมากกว่าบริเวณอื่นในพื้นที่ ในขณะที่ปัญหาที่พบจากการใช้ซอฟต์แวร์คือ จากแบบจำลองต้นไม้การตัดสินใจ จะเรียงลำดับต้นไม้มาเรียงกัน แดกย่อยออกไปตามลำดับของค่าสัดส่วนเกณฑ์ที่มีค่าสูงสุดก่อนแล้วลดหลั่นตามค่าที่น้อยกว่าตามลำดับ ซึ่งไม่ได้ระบุค่าให้ทราบ

ในการวิจัยครั้งต่อไป หากมีการปรับข้อมูลให้มีความละเอียดมากขึ้น (ในการศึกษานี้ข้อมูลที่ใช้ในการวิเคราะห์อ้างอิงบนพื้นที่มาตราส่วนข้อมูลที่ 1:50,000) ซึ่งผลลัพธ์ที่ได้กับการเกิดน้ำท่วมที่ส่งผลต่อความเสียหายทั้งชีวิตและทรัพย์สินจากอุทกภัยนั้นมากมาย มีความต้องการข้อมูลที่มีความถูกต้องและแม่นยำในระดับสูงโดยเฉพาะในเขตชุมชนเมือง ข้อมูลที่มีความละเอียดมากขึ้นจะทำให้การวิเคราะห์มี

ความแม่นยำเชิงตำแหน่งและมีความถูกต้องสูงขึ้นไปด้วย ช่วยให้การวางแผนป้องกันอุทกภัยอย่างเป็นตรงเป้าและลดค่าใช้จ่ายได้ดีมากยิ่งขึ้น ในการศึกษาครั้งนี้ได้นำเสนอวิธีการต้นไม้การตัดสินใจที่นิยมใช้ในงานทางวิทยาการเทคโนโลยีสารสนเทศมาผนวกกับงานทางภูมิสารสนเทศเพื่อให้เป็นแนวทางเลือกทางหนึ่งในการกำหนดพื้นที่เสี่ยงภัยน้ำท่วมเพื่อการหามาตรการรองรับพื้นที่วิกฤตเฉพาะที่มองเห็นภาพได้อย่างชัดเจนช่วยให้ดำเนินการได้อย่างมีประสิทธิภาพในการวางแผนบริหารและจัดการภัยพิบัติอีกแนวทางหนึ่ง

กิตติกรรมประกาศ

ขอขอบคุณสถานวิจัยสารสนเทศภูมิศาสตร์ทรัพยากรธรรมชาติและสิ่งแวดล้อม ศูนย์ภูมิภาคเทคโนโลยีอวกาศและภูมิสารสนเทศ (ภาคใต้) คณะการจัดการสิ่งแวดล้อม มหาวิทยาลัยสงขลานครินทร์ ที่ให้ความอนุเคราะห์ฐานข้อมูลชุดลุ่มน้ำทะเลสาบสงขลา

เอกสารอ้างอิง

1. คณะการจัดการสิ่งแวดล้อมมหาวิทยาลัยสงขลานครินทร์. (ม.ป.ป.). ลักษณะทางชีวภาพ-กายภาพลุ่มน้ำทะเลสาบสงขลา. [Online], 2 ธันวาคม 2559. http://www.songkhulake.com/content/bio_physic
2. สำนักงานป้องกันและบรรเทาสาธารณภัยจังหวัดสงขลา. (ม.ป.ป.). สถิติการเกิดภัยพิบัติ ประจำปี 2555 – 2557. [Online], 18 ธันวาคม 2559. http://122.155.1.145/inner.dpmsk-1.146/cms/inner_758/2927.1/
3. นาถนเรศ อาภาสุวรรณ, ประมาณ เทพสงเคราะห์ และ วรุฒม์ นาทิ. การศึกษาปัจจัยการเกิดน้ำท่วมเพื่อกำหนดพื้นที่เสี่ยงภัยน้ำท่วมด้วยระบบสารสนเทศภูมิศาสตร์และแนวทางป้องกันบรรเทาในบริเวณลุ่มน้ำย่อย ทะเลสาบสงขลาฝั่งตะวันตก จังหวัดพัทลุง. วารสารมนุษยศาสตร์สังคมศาสตร์ มหาวิทยาลัยทักษิณ. 2551; 3(2): 176-199.
4. กรมป้องกันและบรรเทาสาธารณภัย กระทรวงมหาดไทย. (8 ธันวาคม 2559). ปภ.รายงานสถานการณ์อุทกภัยในพื้นที่ภาคใต้คลี่คลายแล้ว 4 จังหวัด ยังคงมีสถานการณ์ใน 8 จังหวัด.[Online], 10 มกราคม 2560. http://www.disaster.go.th/th/cdetail-7927-mission_managers-119-1/
5. สุพิชฌาย์ ธนารุณ. การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการกำหนดพื้นที่เสี่ยงอุทกภัยจังหวัดอ่างทอง, คณะพัฒนาสังคมและสิ่งแวดล้อม สถาบันบัณฑิตพัฒนบริหารศาสตร์. 2553.
6. สุนธุ์พงศ์ สูงสุมาลย์. การใช้ระบบสารสนเทศภูมิศาสตร์เพื่อศึกษาสภาพ และสาเหตุผลกระทบจากอุทกภัยในพื้นที่อำเภอประจันตคาม, บัณฑิตศึกษา เทคโนโลยีระบบสารสนเทศภูมิศาสตร์ มหาวิทยาลัยเทคโนโลยีมหานคร. 2549.
7. พีระพิทย์ พีชมงคล และคณะ. การกำหนดเขตพื้นที่น้ำท่วมในภาคใต้ฝั่งตะวันออกตอนล่างของประเทศไทยโดยใช้ระบบภูมิสารสนเทศ, วารสารสมาคมสำรวจข้อมูลระยะไกลและสารสนเทศภูมิศาสตร์แห่งประเทศไทย. 2550; 8(3): 47-57.
8. ศูนย์ภูมิภาค สำนักพัฒนาอุตุนิยมวิทยา กรมอุตุนิยมวิทยา. (ม.ป.ป.). ภูมิภาคของประเทศไทย. [Online], 27 สิงหาคม 2558. https://www.tmd.go.th/info/climate_of_thailand-2524-2553.pdf
9. กรมพัฒนาที่ดิน กระทรวงเกษตรและสหกรณ์. คู่มือการเขียนหน่วยแผนที่ดิน ส่วนมาตรฐานการสำรวจจำแนกดินและที่ดิน สำนักสำรวจดินและวางแผนการใช้ที่ดิน. เอกสารวิชาการฉบับที่ 519 กรกฎาคม 2547.
10. ส่วนวิเคราะห์สภาพการใช้ที่ดิน สำนักสำรวจดินและวางแผนการใช้ที่ดิน กรมพัฒนาที่ดิน. การจำแนกใช้การประโยชน์ที่ดิน (Land Use Classification). กรุงเทพฯ: กรมพัฒนาที่ดิน 2552.
11. จิตตฤ พูลวัน. ระบบวิเคราะห์โรคทั่วไปโดยใช้การจำแนกข้อมูลบนอินเทอร์เนต ภาควิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2550.
12. ขจรศักดิ์ ศรีอ่อน. การทำนายสาเหตุของเหตุการณ์กระแสไฟฟ้าขัดข้อง โดยใช้เทคนิคการทำเหมืองข้อมูลในระบบจำหน่ายของการไฟฟ้าส่วนภูมิภาค เขต 1 ภาคกลาง. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมไฟฟ้า) สาขาวิชาวิศวกรรมไฟฟ้าภาควิชาวิศวกรรมไฟฟ้าบัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์. 2552.
13. Geo-Informatics and Space Technology Development Agency (Public Organization). Thailand Flood Monitoring System. [Online]. Available from: <http://flood.gistda.or.th/>. Accessed January 16, 2016.
14. Han, J. and M. Kamber. Data Mining: Concepts and Techniques. CA: Morgan Kaufmann, San Francisco. 2001.
15. พยูน พาณิชย์กุล. การพัฒนาระบบดัดจำไม้หนึ่งโดยใช้ Decision Tree. โครงการพัฒนาระบบงานปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนง

- วิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าคุณ
ทหารลาดกระบัง. 2548.
16. Quinlan, J. R. C4.5: program for machine learning.
Morgan Kaufmann. 1992.
 17. อนันต์ ปิ่นะเต. การพัฒนาระบบสนับสนุนการตัดสินใจใน
การเลือกสมัครในสาขาวิชา โดยใช้เทคนิคต้นไม้ตัดสินใจ.
วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม.
2559; 35(4): 413-421.
 18. Pontius, R. G. & Schneider, L. C. Land-use change
model validation by an ROC method for the Ipswich
watershed, Massachusetts, USA. Agriculture, Eco-
systems and Environment. 2001; 85: 239-248.
 19. Thomas, G. & Tape, M.D. The Area Under an ROC
Curve. University of Nebraska Medical Center. [On-
line]. Available from: [http://gim.unmc.edu/dxtests/
roc3.htm](http://gim.unmc.edu/dxtests/roc3.htm) Accessed September 26, 2016.