

ระบบคัดกรองโรคมะเร็งปากมดลูก โดยใช้เทคนิคการทำเหมืองข้อมูล  
แบบการจำแนกประเภทข้อมูล  
Cervical cancer screening system by using a data mining  
technique Classification

ปิยดา เกลี้ยงจิตร์<sup>1</sup>, สุภัทสรุ ขาวปาน<sup>1</sup>, วิภาวรรณ บัวทอง<sup>1</sup> และ ณสิทธิ์ เหล่าเส็น<sup>1</sup>

**บทคัดย่อ**

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความเสี่ยงโรคมะเร็งปากมดลูก ซึ่งวิเคราะห์ข้อมูลจากผู้ป่วยโรคมะเร็งปากมดลูกในโรงพยาบาล Universitario de Caracas จำนวน 858 คน 36 แอตทริบิวต์ อาศัยข้อมูล 8 ด้าน คือ Age (อายุ), Smokes (สูบบุหรี่), Dx (การวินิจฉัยโรค), Schiller (วินิจฉัยทางเคมี), Cytology (วินิจฉัยทางช่องท้องหรือปัสสาวะ), Dx : Cancer (การตรวจหาโรคมะเร็ง), Biopsy (การวินิจฉัยตรวจชิ้นเนื้อ) และ Risk (ความเสี่ยง) ด้วยเทคนิคการทำเหมืองข้อมูลรูปแบบการจำแนกมาวิเคราะห์ที่มีความเสี่ยงของผู้ป่วยโรคมะเร็งปากมดลูก การใช้อัลกอริทึม 3 รูปแบบ ได้แก่ REP TREE, C4.5 และ IBK ผลการวิจัยพบว่า อัลกอริทึม REP TREE มีความแม่นยำมากที่สุด โดยมีวิธีการประเมินประสิทธิภาพของโมเดล คือ Cross-validation 10 Folds ซึ่งให้ค่า Accuracy 0.989, Precision 0.988, Recall 1.00 และค่า F - measure 0.994 เป็นเงื่อนไขที่แม่นยำและน่าเชื่อถือมากที่สุด จึงนำข้อมูลการวิจัยมาใช้ในการตรวจคัดกรองโดยใช้เทคนิคการจำแนกข้อมูลผู้ที่มีความเสี่ยงต่อมะเร็งปากมดลูก เพื่อใช้ในการนำมาเขียนเว็บแอปพลิเคชันสำหรับผู้ใช้งานทั่วไปที่ต้องการศึกษาความรู้และความเสี่ยงของโรคมะเร็งปากมดลูก

**คำสำคัญ:** โรคมะเร็งปากมดลูก, การทำเหมืองข้อมูล, การจำแนกประเภทข้อมูล

**Abstract**

The objective of this research was to analyze the factors contributing to the risk of cervical cancer. Which analyzes data from patients with cervical cancer in the hospital Universitario de Caracas 858 people, 36 attributes based on 8 data areas: Age (age), Smokes (smoking), Dx (diagnosis), Schiller (chemical diagnosis), Cytology (abdominal or urinary diagnosis), Dx: Cancer (cancer detection), Biopsy (biopsy diagnosis) and Risk (risk) by data mining techniques, classification models, analysis at risk of patients with cervical cancer Using three algorithms: REP TREE, C4.5 and IBK, the research found that the REP TREE algorithm was the most accurate. The method for evaluating the model's performance is Cross-validation 10 Folds, which gives Accuracy 0.989, Precision 0.988, Recall 1.00 and F - measure 0.994 is the most accurate and reliable condition. Therefore, the research data was used for screening using the technique of classification of people at risk of cervical cancer. For use in writing web applications for general users who want to study the knowledge and risks of cervical cancer.

**Keywords:** Cervical Cancer, Data Mining, Classification

<sup>1</sup> คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต

## บทนำ

โรคมะเร็งปากมดลูก (Cervical Cancer) เป็นปัญหาทางสาธารณสุขที่สำคัญของประเทศไทย โดยพบเป็นอันดับ 2 ของมะเร็งในสตรีไทย รองจากมะเร็งเต้านม จากรายงานของ WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre), Human Papillomavirus and Related Cancers in Thailand, Summary Report 2010 ได้รายงานในปี พ.ศ.2551 ประเทศไทยมีประชากรสตรีที่เสี่ยงต่อมะเร็งปากมดลูก 26.09 ล้านคน มีอายุตั้งแต่ 15 ปี ขึ้นไป มีจำนวนผู้ป่วยมะเร็งปากมดลูกรายใหม่ปีละ 9,999 ราย เสียชีวิต 5,216 ราย หรือประมาณร้อยละ 53 ถ้าคิดเป็นวันแล้วจะมีสตรีไทยเสียชีวิตจากมะเร็งปากมดลูกเฉลี่ยวันละ 14 คน อุบัติการณ์โดยรวมของมะเร็งปากมดลูกในช่วง 15 ปีที่ผ่านมาลดลงเพียงเล็กน้อย แต่ในจำนวนผู้ป่วยรายใหม่สูงขึ้นเรื่อย ๆ ถ้าอุบัติการณ์ของมะเร็งปากมดลูกยังคงสูงในระดับนี้โดยไม่มีมาตรการหรือยุทธวิธีใหม่ที่มีประสิทธิภาพมาป้องกัน คาดว่าในปี พ.ศ.2568 จะมีจำนวนผู้ป่วยรายใหม่สูงขึ้นเป็นปีละประมาณ 13,082 คน เสียชีวิต 7,871 ราย สาเหตุของการเกิดโรคมะเร็งปากมดลูกยังไม่ทราบแน่ชัด (ไพฑูริย์ ออบเชย,2553)

ปัจจุบันการดำเนินงานควบคุมป้องกันมะเร็งปากมดลูกของประเทศไทยคือ การตรวจคัดกรองเพื่อค้นหาความผิดปกติหรือรอยโรคที่ปากมดลูกให้พบก่อนที่จะกลายเป็นมะเร็ง วิธีการตรวจคัดกรองที่นิยมใช้มี 2 วิธี คือ การตรวจแปปสเมียร์ (Pap Smear) และการตรวจโดยใช้น้ำส้มสายชู หรือ Visual Inspection with Acetic acid (VIA) จากการศึกษาของนักวิจัยโครงการประเมินเทคโนโลยีและนโยบายด้านสุขภาพ (HITAP) พบว่าการตรวจคัดกรองมะเร็งปากมดลูกด้วยวิธีแปปสเมียร์ทุก 5 ปี เป็นการลงทุนที่คุ้มค่าเมื่อเทียบกับการไม่ได้ตรวจคัดกรองโรคเลย และต้นทุนในการตรวจคัดกรองโรคมะเร็งปากมดลูกมีมูลค่าต่ำกว่าต้นทุนที่ต้องใช้ในการรักษามะเร็งปากมดลูกในกรณีที่ไม่มี การตรวจคัดกรองและ International Agency for Research on Cancer (IARC) ได้ทำการศึกษพบว่าหากตรวจพบผู้ป่วยมะเร็งปากมดลูกในระยะที่หนึ่งอัตราการรอดชีพ 5 ปี จะสูงถึงร้อยละ 90-95 แต่หากตรวจพบผู้ป่วยในระยะที่สี่ซึ่งเป็นระยะสุดท้าย อัตราการรอดชีพ 5 ปี จะลดลงเหลือเพียงน้อยกว่าร้อยละ 5 นอกจากนี้ IARC/World Health Organization (WHO) ยังทำการศึกษพบว่าหากทำแปปสเมียร์ 1 ครั้ง ทุกปี, 1 ครั้งทุก 2 ปี หรือ 1 ครั้งทุก 3 ปี อัตราการเกิดมะเร็งปากมดลูกจะลดลงร้อยละ 91-93 หากทำแปปสเมียร์ 1 ครั้ง ทุก 5 ปี อัตราการเกิดมะเร็งปากมดลูกจะลดลง ร้อยละ 84 (สุตาฟ้า วงศ์หิรามาศย์,2556)

จากการศึกษางานวิจัยการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนโดยใช้ข้อมูลของประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ปี พ.ศ. 2555 จำนวน 1,071 เรคอร์ด ด้วยเทคนิคการทำเหมืองข้อมูล โดยแบ่งการศึกษาเป็น 2 ประเด็นคือ การศึกษาความสัมพันธ์ของการเป็นโรคต่าง ๆ โดยใช้เทคนิค Association Rule และคัดกรองสุขภาพเบื้องต้นของประชาชนโดยแบ่งออกเป็น กลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย ด้วยเทคนิค Classification และนำผลลัพธ์ที่ได้ไปพัฒนาเป็นระบบคัดกรองสุขภาพเบื้องต้น ผลการทดลองพบว่า อัลกอริทึม FP Growth ให้คำตอบที่เหมาะสมมากกว่าอัลกอริทึม Apriori และในการศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชน พบว่าอัลกอริทึม Partial Rules ให้ความถูกต้องมากกว่าอัลกอริทึมต้นไม้ตัดสินใจ C4.5 โดยมีค่าความถูกต้อง (precision) เท่ากับ 88.60% ค่าระลึก (recall) เท่ากับ 89.20% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.80% ได้ผลลัพธ์เป็นกฎจำนวน 42 กฎ และถูกนำไปใช้ในการสร้างระบบคัดกรองสุขภาพอัตโนมัติสำหรับให้ประชาชนทั่วไปเข้าตรวจสอบสุขภาพเบื้องต้นของตนเอง (กิตติศักดิ์ สุมาลย์,2555) การประยุกต์ใช้เทคนิคจำแนกข้อมูลแบบต้นไม้เพื่อการวินิจฉัยโรคในโคเบื้องต้นบนโทรศัพท์มือถือ ผู้วิจัยได้คัดเลือกปัจจัยที่เกี่ยวข้องสำหรับการวินิจฉัยโรคได้ทั้งหมด 22 ปัจจัย และสร้างโมเดลโดยใช้เทคนิคการจำแนกประเภทข้อมูลด้วยวิธีต้นไม้ตัดสินใจใช้อัลกอริทึมจำนวน 3 ตัว ได้แก่ J48, Random Tree และ REP Tree แล้วทำการทดสอบตัวแบบการพยากรณ์ด้วยวิธีการตรวจสอบแบบไขว้ เพื่อที่จะหาโมเดลที่ได้ไปพัฒนาแอปพลิเคชัน ผลการวิจัยพบว่า ต้นไม้ตัดสินใจแบบอัลกอริทึม Random Tree ที่การแบ่งทดสอบแบบ 2-fold cross validation มีประสิทธิภาพค่าความถูกต้องสูงสุดที่สุด มีค่าเท่ากับ 99.47% และมีการพยากรณ์ที่เที่ยงตรงจากการพิจารณาค่าความคลาดเคลื่อนระหว่างค่าจริงและค่าที่พยากรณ์ได้เท่ากับ 0.020 ค่าความแม่นยำเท่ากับ 0.995 ค่าความระลึกเท่ากับ 0.995 และค่าความถ่วงดุลเท่ากับ 0.995 จากนั้นนำอัลกอริทึมที่ให้ค่าความแม่นยำสูงสุดไปพัฒนาเป็นแอปพลิเคชันบนโทรศัพท์มือถือ (ณัฐวดี หงษ์บุญมีและพงศันรินทร์ ศรีรุ่ง,2561)

การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล ผู้วิจัยเก็บรวบรวมข้อมูลในสาขาแพทย์และชีวการแพทย์จากฐานข้อมูล UCI ในเว็บไซต์ที่ได้มีการรวบรวมไว้แล้วจำนวน 3 ชุด ได้แก่ ชุดข้อมูล fertility ขนาดตัวอย่างที่ใช้ คือ 100 หน่วยตัวอย่าง ชุดข้อมูล vertebral column ขนาดตัวอย่างที่ใช้ คือ 310 หน่วยตัวอย่างและชุดข้อมูล diabetes ขนาดตัวอย่างที่ใช้ คือ 768 หน่วยตัวอย่าง โดยวิธีการจำแนกที่นำมา

เปรียบเทียบมี 7 วิธี ได้แก่ วิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้อัลกอริทึมชนิด IBK วิธีต้นไม้ตัดสินใจโดยใช้อัลกอริทึม SMO ชนิด โพลีโนเมียลเคอร์เนล วิธีฐานกฎโดยใช้อัลกอริทึม decision table วิธีการถดถอยโลจิสติกส์ทวิภาค และวิธีนาอ็อบเบย์ การเปรียบเทียบผลการศึกษาพบว่าชุดข้อมูล fertility วิธีการถดถอยโลจิสติกส์ทวิภาคที่ random seed = 10, 20 และ 30 มีความถูกต้อง ค่าความไว ค่าความจำเพาะ และค่าคลาดเคลื่อนกำลังสองเฉลี่ยดีที่สุดคือ ร้อยละ 100, 1.000, 1.0000 และ 0.00000 ตามลำดับ ส่วนชุดข้อมูล vertebral volume วิธีเพื่อนบ้านใกล้สุด k ตัว ที่ random seed = 10, 20 และ 30 มีความถูกต้อง ค่าความไว ค่าความจำเพาะ และค่าคลาดเคลื่อนกำลังสองเฉลี่ยดีที่สุดคือ ร้อยละ 100, 1.0000, 1.0000 และ 0.00024 ตามลำดับ และชุดข้อมูล diabetes วิธีเพื่อนบ้านใกล้สุด k ตัว ที่ random seed = 10, 20 และ 30 มีความถูกต้อง ค่าความไว ค่าความจำเพาะ และค่าคลาดเคลื่อนกำลังสองเฉลี่ยดีที่สุดคือ ร้อยละ 100, 1.0000, 1.0000 และ 0.00004 ตามลำดับ และเมื่อพิจารณาข้อมูลทั้ง 3 ชุด ร่วมกับ พบว่าวิธีที่ดีที่สุดในการทำนายผล คือ วิธีเพื่อนบ้านใกล้สุดโดยใช้อัลกอริทึมชนิด IBK (สายชล สินสมบุรณ์ทอง,2563) ซึ่งงานวิจัยทั้งหมดนี้เป็นแนวคิดที่สามารถนำมาเพื่อวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความเสี่ยงโรคมะเร็งปากมดลูก เพื่อให้ผู้ใช้งานทราบถึงสาเหตุการก่อให้เกิดโรคมะเร็งปากมดลูกและเพื่อให้ความรู้เกี่ยวกับโรคมะเร็งปากมดลูกและวิธีป้องกัน โดยข้อมูลที่ใช้มาจากเว็บไซต์ UCI โรงพยาบาล Universitario de Caracas เมืองการากัส ประเทศเวเนซุเอลา (Kelwin Fernandes, Jaime S. Cardoso, and Jessica Fernandes,2017) จำนวน 858 คน 36 แอทริบิวต์

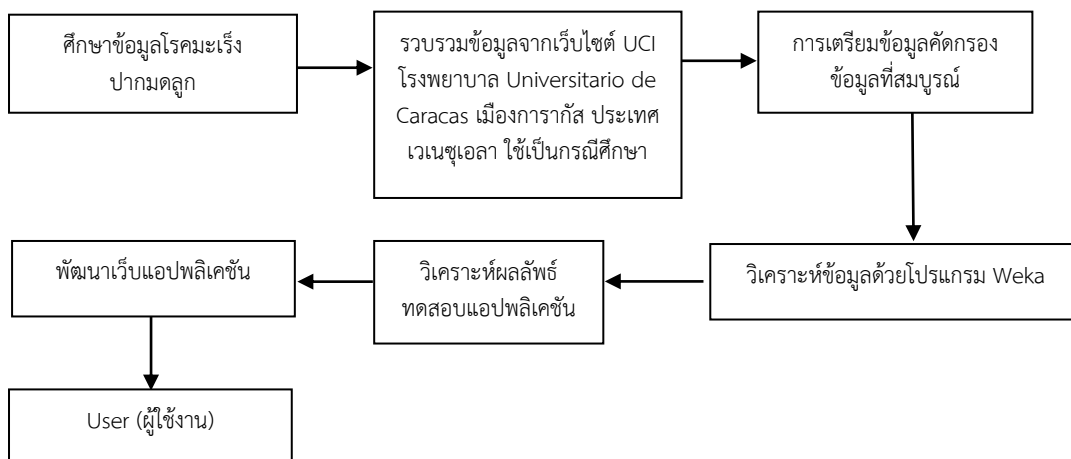
### วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความเสี่ยงโรคมะเร็งปากมดลูก
2. เพื่อให้ผู้ใช้งานทราบถึงสาเหตุการก่อให้เกิดโรคมะเร็งปากมดลูก
3. เพื่อให้ความรู้เกี่ยวกับโรคมะเร็งปากมดลูกและวิธีป้องกัน

### วิธีดำเนินการวิจัย

องค์ประกอบที่เกี่ยวข้องในขั้นตอนการศึกษาเพื่อวิเคราะห์ข้อมูลของผู้ป่วยที่มีความเสี่ยงต่อโรคมะเร็งปากมดลูกมีดังนี้

1. รวบรวมข้อมูลและเตรียมข้อมูล
2. สร้างแบบจำลองการจำแนกความเสี่ยงโรคมะเร็งปากมดลูก
3. พัฒนาแอปพลิเคชัน โดยใช้แบบจำลองการจำแนก



ภาพที่ 1 ขั้นตอนการศึกษาและพัฒนาเว็บแอปพลิเคชัน

การประชุมวิชาเทคโนโลยีอุตสาหกรรม ระดับชาติ ครั้งที่ 2  
 ในวาระครบ 65 ปี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

1. ศึกษาข้อมูลโรคมะเร็งปากมดลูก ทั้งสาเหตุการเกิดโรค อาการ และวิธีการป้องกันโรค
2. รวบรวมข้อมูลและเตรียมข้อมูล ค้นหาข้อมูลของโรคมะเร็งปากมดลูกจากเว็บไซต์ UCI และรวบรวมข้อมูลผู้ป่วยโรคมะเร็งปากมดลูก ในปี 2560 ของโรงพยาบาล Universitario de Caracas เมืองการากัส ประเทศเวเนซุเอลา โดยมีจำนวนทั้งหมด 858 คน 36 แอตทริบิวต์ นำข้อมูลมาทำการเตรียมข้อมูลที่สามารถเข้าไปสู่วิธีการทำเหมืองข้อมูลต่อไปได้
3. สร้างแบบจำลองการจำแนกความเสี่ยงโรคมะเร็งปากมดลูก ผู้วิจัยนำข้อมูลมาวิเคราะห์คัดกรองจำแนกความเสี่ยงของโรคมะเร็งปากมดลูก โดยใช้เทคนิคเหมืองข้อมูล Classification และทำการวิเคราะห์โดยใช้อัลกอริทึม ได้แก่ IBK,C4.5 และ REP TREE ผลลัพธ์การวิเคราะห์จากอัลกอริทึมที่มีความแม่นยำและดีที่สุด คือ REP TREE อาศัยค่า Accuracy, Precision, Recall และ F-Measure
4. พัฒนาเว็บแอปพลิเคชัน นำข้อมูลแบบจำลองการจำแนกในข้อที่ 3 มาพัฒนาเว็บแอปพลิเคชันระบบคัดกรองความเสี่ยงโรคมะเร็งปากมดลูก และเครื่องมือที่ใช้ได้แก่ ionic framework, Visual Studio Code, Node.js เว็บแอปพลิเคชันมีทั้งหมด 5 page

ผลการศึกษา

จากแหล่งข้อมูล เว็บไซต์ UCI cancer cervical

1	Age	Number of First sexua	Num of pri	Smokes (y	Smokes (p
2	18	4	15	1	0
3	15	1	14	1	0
4	34	1 ?		1	0
5	52	5	16	4	1
6	46	3	21	4	0
7	42	3	23	2	0
8	51	3	17	6	1
9	26	1	26	3	0
10	45	1	20	5	0
11	44	3	15 ?		1
12	44	3	26	4	0
13	27	1	17	3	0
14	45	4	14	6	0
15	44	2	25	2	0
16	43	2	18	5	0
17	40	3	18	2	0
18	41	4	21	3	0
19	43	3	15	8	0
20	42	2	20 ?		0
21	40	2	27 ?		0
22	43	2	18	4	0
23	41	3	17	4	0
24	40	1	18	1	0
25	40	1	20	2	0
26	40	3	15	3	0
27	44	3	19	1	0
28	39	5	23	2	0
29	39	2	17	4	0
30	37	3	24	1	1
31	37	3	26	1	0

ภาพที่ 2 ข้อมูลที่ยังไม่ผ่าน Data Cleaning มีจำนวน 858 คน 36 แอตทริบิวต์ ข้อมูลของผู้ป่วยไม่ชัดเจน แสดงเป็นเครื่องหมาย ? ไม่สามารถบอกจำนวนหรือตัวเลขได้ชัดเจน

การประชุมวิชาการเทคโนโลยีอุตสาหกรรม ระดับชาติ ครั้งที่ 2  
 ในวาระครบ 65 ปี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

19: Dx:Cancer	20: Dx:CN	21: Dx:HPV	22: Dx:	23: Hinselmann	24: Schiller	25: Cytology	26: Biopsy	27: Risk
Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	Yes
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	Yes
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	Yes
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	Yes
1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	Yes
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	No

ภาพที่ 3 ข้อมูลที่ทำการ Data Cleaning แล้วแปลงเป็นไฟล์ .csv

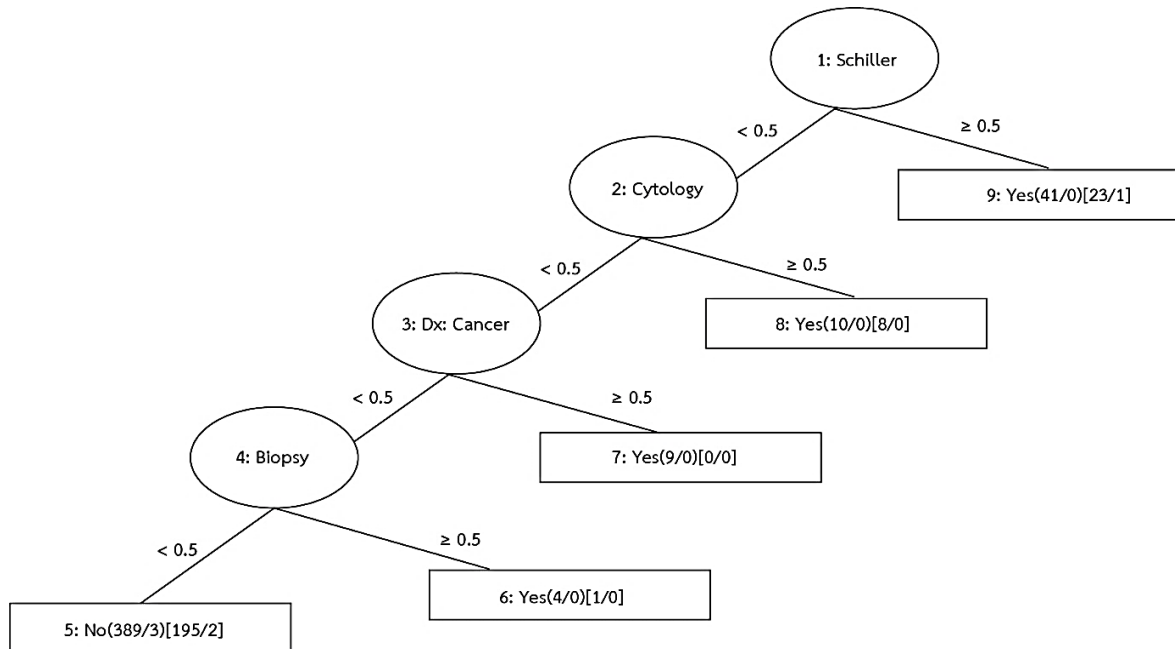
จากข้อมูลเดิมมีจำนวน 858 คน 36 แอดทริบิวต์ ได้ทำการคัดกรองข้อมูลบางส่วนที่ไม่สมบูรณ์ ข้อมูลแอดทริบิวต์ที่ซ้ำกัน จึงได้จำนวนผู้ป่วยและแอดทริบิวต์ที่มีข้อมูลที่สมบูรณ์ 679 คน 27 แอดทริบิวต์ ดังนี้

ตารางที่ 1 แอดทริบิวต์ 27 แอดทริบิวต์

Attribute	ความหมาย	ชนิดของข้อมูล
Age	อายุ	Numeric
Number of Sexual partners	จำนวนคู่นอน	Numeric
First Sexual intercourse	การมีเพศสัมพันธ์ครั้งแรก	Numeric
Num of pregnancies	จำนวนการมีบุตร	Numeric
Smokes	สูบบุหรี่	Numeric
STDs	โรคติดต่อทางเพศสัมพันธ์ (กามโรค)	Numeric
STDs: cervical condylomatosis	หูดหรือโรคอวัยวะเพศ	Numeric
STDs: vaginal condylomatosis	การติดเชื้อ	Numeric
STDs: vulvo-perineal condylomatosis	การรักษาฝีเย็บ	Numeric
STDs: syphilis	โรคติดเชื้อจากแบคทีเรีย	Numeric
STDs: pelvic inflammatory disease	ภาวะอักเสบในอุ้งเชิงกราน	Numeric
STDs: genital herpes	โรคเริม	Numeric
STDs: molluscum contagiosum	โรคหูดข้าวสุก	Numeric
STDs: AIDs	โรคเอดส์	Numeric
STDs: HIV	ติดเชื้อเอชไอวี	Numeric
STDs: Hepatitis B	ติดเชื้อไวรัสตับอักเสบบี	Numeric
STDs: HPV	การติดเชื้อเนื้อเยื่อปาก	Numeric
STDs: Number of diagnosis	การวินิจฉัยโรคร่วม	Numeric
STDs: Cancer	โรคมะเร็ง	Numeric
Dx: Cancer	การวินิจฉัยโรคมะเร็ง	Numeric
Dx: HPV	วินิจฉัยการติดเชื้อเนื้อเยื่อปาก	Numeric
Dx:	การวินิจฉัยโรค	Numeric
Hinselmann	การส่องกล้องปากมดลูกด้วยคอลโปสโคป	Numeric
Schiller	การวินิจฉัยทางเคมี	Numeric
Cytology	การวินิจฉัยทางช่องท้องหรือปัสสาวะ	Numeric
Biopsy	การวินิจฉัยตรวจชิ้นเนื้อ	Numeric
Risk	ความเสี่ยง	Class

การประชุมวิชาเทคโนโลยีอุตสาหกรรม ระดับชาติ ครั้งที่ 2  
 ในวาระครบ 65 ปี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

ตารางที่ 1 แอตทริบิวต์ข้อมูลแบบสอบถามของผู้ใช้งานในการประเมินความเสี่ยง พฤติกรรมที่ก่อให้เกิดความเสี่ยง โรคมะเร็งปากมดลูก โรคที่เคยเกิดขึ้นกับผู้ใช้งานและการวินิจฉัยโรค



ภาพที่ 4 แผนภาพการจำแนกข้อมูลแบบ Visualizer Tree

จากภาพที่ 4 สามารถแปลผลได้ดังนี้ มี 5 ข้อ

- กฎข้อที่ 1 ผู้ใช้งานเคยมีการวินิจฉัยทางเคมี แต่มีการวินิจฉัยทางช่องท้องหรือปัสสาวะ มีความเสี่ยงสูงที่จะเป็นโรคมะเร็งปากมดลูก
- กฎข้อที่ 2 ผู้ใช้งานไม่เคยมีการวินิจฉัยทางเคมี แต่มีการวินิจฉัยทางช่องท้องหรือปัสสาวะ มีความเสี่ยงสูงที่จะเป็นมะเร็งปากมดลูก
- กฎข้อที่ 3 ผู้ใช้งานไม่เคยมีการวินิจฉัยทางเคมีและวินิจฉัยทางช่องท้องหรือปัสสาวะ แต่มีการวินิจฉัยโรคมะเร็ง มีความเสี่ยงสูงที่จะเป็นมะเร็งปากมดลูก
- กฎข้อที่ 4 ผู้ใช้งานไม่เคยมีการวินิจฉัยทางเคมี การวินิจฉัยทางช่องท้องหรือปัสสาวะและการวินิจฉัยโรคมะเร็ง แต่มีการวินิจฉัยตรวจชิ้นเนื้อ มีความเสี่ยงสูงที่จะเป็นโรคมะเร็งปากมดลูก
- กฎข้อที่ 5 ผู้ใช้งานไม่เคยมีการวินิจฉัยตรวจชิ้นเนื้อ มีความเสี่ยงน้อยที่จะเป็นมะเร็งปากมดลูก

ตารางที่ 2 สรุปผลการทดลองอัลกอริทึม ดังนี้

Algorithm	Cross-validation Folds 5				Cross-validation Folds 10				Percentage split % 10			
	Accuracy	Precision	Recall	F-Measure	Accuracy	Precision	Recall	F-Measure	Accuracy	Precision	Recall	F-Measure
REP TREE	0.986	0.985	1.00	0.992	0.989	0.988	1.00	0.994	0.947	0.942	1.00	0.970
IBK	0.985	0.983	1.00	0.991	0.985	0.983	1.00	0.991	0.975	0.972	1.00	0.986
C4.5	0.949	0.945	1.00	0.945	0.952	0.948	1.00	0.973	0.947	0.942	1.00	0.970

ตารางที่ 2 จากการทดลองด้วยโปรแกรม Weka โดยใช้การจำแนก Data Classification ได้แก่อัลกอริทึม IBK, C4.5 และ REP TREE โดยใช้วิธี Cross-validation Folds5 ,Folds 10 และ Percentage split % 10 อาศัย Accuracy, Precision, Recall และ F-Measure เพื่อหาความถูกต้องแม่นยำของ Model แสดงผลได้ดังตารางที่ 1 โดยใช้วิธี 2 วิธี ได้แก่ 1) Cross-validation Test อัลกอริทึมที่ได้ค่ามากที่สุด คือ REP TREE folds 10 แสดงค่า Accuracy 0.989 ,Precision 0.985,Recall 1.00 และ F-Measure 0.994 อัลกอริทึมที่มีค่าต่ำสุด คือ C4.5 folds 5 แสดงค่า Accuracy 0.949,Precision 0.945,Recall 1.00 และ F-Measure 0.971 2) Percentage split % 10 อัลกอริทึมที่ได้ค่ามากที่สุดคือ IBK มีค่า Accuracy 0.975,Precision 0.972,Recall 1.00 และ F-Measure 0.986 และอัลกอริทึมที่ได้ค่าต่ำสุดมี 2 อัลกอริทึมค่าเท่ากัน ได้แก่ REP TREE และ C4.5 แสดงค่า Accuracy 0.947, Precision 0.942,Recall 1.00 และ F-Measure 0.970

## 2. ผลการพัฒนาเว็บแอปพลิเคชัน

จากภาพที่ 4 แสดงผลข้อมูลการจำแนกที่ดีที่สุดเป็นไปตามเงื่อนไขที่ตั้งไว้ ดังนั้น ผู้วิจัยจึงพัฒนาเว็บแอปพลิเคชัน โดยการเขียนโปรแกรมให้อยู่ในรูปแบบเว็บแอปพลิเคชัน มีทั้งหมด 5 page ดังภาพที่ 5 - 9

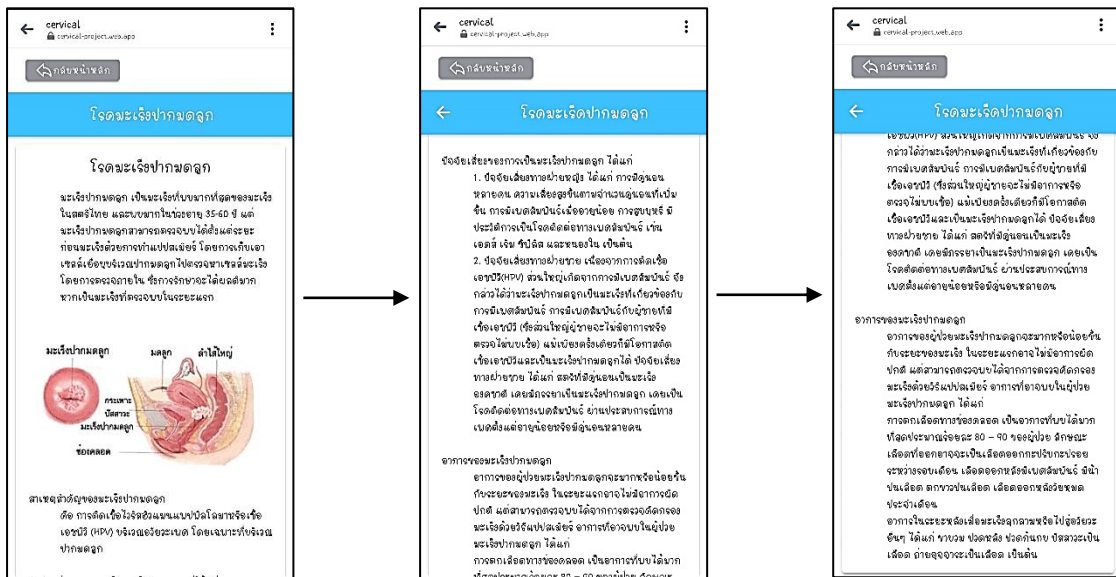


ภาพที่ 5 หน้าแรกของเว็บแอปพลิเคชัน ผู้ใช้สามารถใส่ชื่อของตนเองได้ แล้วคลิก Login เพื่อเข้าสู่หน้าเมนู

การประชุมวิชาเทคโนโลยีอุตสาหกรรม ระดับชาติ ครั้งที่ 2  
 ในวาระครบ 65 ปี มหาวิทยาลัยราชภัฏนครศรีธรรมราช



ภาพที่ 6 เมนูหลัก 3 เมนู ได้แก่ ความรู้เกี่ยวกับโรคมะเร็งปากมดลูก วิธีป้องกัน และตรวจประเมินความเสี่ยง

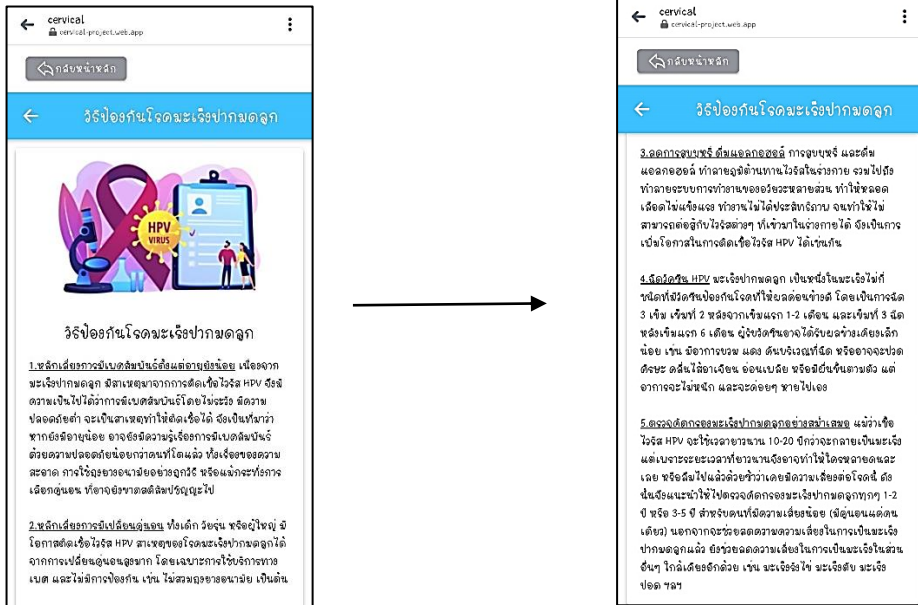


ภาพที่ 7 ความรู้เกี่ยวกับโรคมะเร็งปากมดลูก

เมนูหลักที่ 1 ความรู้เกี่ยวกับโรคมะเร็งปากมดลูก มี 3 หัวข้อหลัก 1) สาเหตุสำคัญของมะเร็งปากมดลูก 2) ปัจจัยเสี่ยงของการเป็นมะเร็งปากมดลูก 3) อาการของมะเร็งปากมดลูก (ธัญพัชร ดีทองอ่อน, 2552)

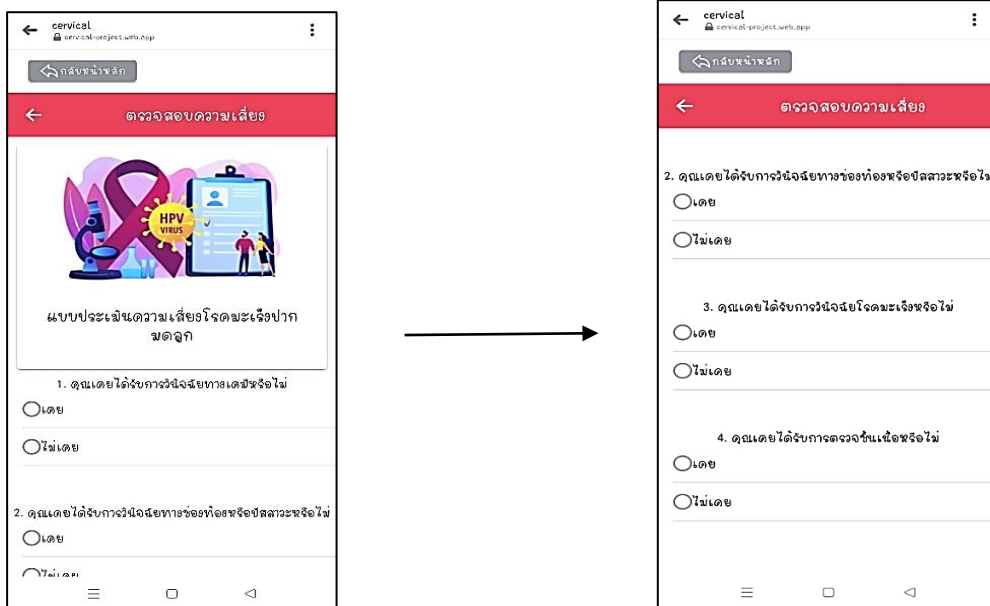


การประชุมวิชาเทคโนโลยีอุตสาหกรรม ระดับชาติ ครั้งที่ 2  
 ในวาระครบ 65 ปี มหาวิทยาลัยราชภัฏนครศรีธรรมราช



ภาพที่ 8 วิธีป้องกันโรคมะเร็งปากมดลูก

เมนูหลักที่ 2 วิธีป้องกันโรคมะเร็งปากมดลูก มีหัวข้อย่อย 5 หัวข้อ ดังนี้ 1) หลีกเลี่ยงการมีเพศสัมพันธ์ตั้งแต่อายุยังน้อย 2) หลีกเลี่ยงการเปลี่ยนคู่นอน 3) ลดการสูบบุหรี่ ตี๋มแอลกอฮอล์ 4) ฉีดวัคซีน HPV และ 5) ตรวจคัดกรองมะเร็งปากมดลูกอย่างสม่ำเสมอ (จตุพล ศรีสมบุญ, 2555)



ภาพที่ 9 ตรวจประเมินความเสี่ยงโรคมะเร็งปากมดลูก

เมนูหลักที่ 3 ตรวจประเมินความเสี่ยง มี 4 คำถาม ดังนี้

- 1) คุณเคยได้รับการวินิจฉัยทางเคมีหรือไม่
- 2) คุณเคยได้รับการวินิจฉัยทางช่องท้องหรือปัสสาวะหรือไม่

การประชุมวิชาเทคโนโลยีอุตสาหกรรม ระดับชาติ ครั้งที่ 2  
 ในวาระครบ 65 ปี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

- 3) คุณเคยได้รับการวินิจฉัยโรคมะเร็งหรือไม่
- 4) คุณเคยได้รับการตรวจชิ้นเนื้อหรือไม่

จากการทำแบบจำลองโมเดล นำมาสร้างเป็นเว็บแอปพลิเคชัน ผู้วิจัยทำการประเมินแบบสอบถามความพึงพอใจจากผู้ใช้งาน จำนวน 20 คน ได้ผลลัพธ์ดังต่อไปนี้

ตารางที่ 3 ผลการประเมินความพึงพอใจจากผู้ทดลองใช้งานเว็บแอปพลิเคชัน

ประเด็น	ค่าคะแนนเฉลี่ย	เกณฑ์ความพึงพอใจ
เนื้อหาที่น่าสนใจ	4.35	มากที่สุด
เว็บแอปพลิเคชัน Cervical Cancer ใช้งานง่าย	4.30	มากที่สุด
เว็บแอปพลิเคชัน Cervical Cancer มีการใช้ตัวอักษรอ่านง่าย น่าสนใจ	4.18	มาก
ได้รับความรู้ในวิธีการป้องกันโรคมะเร็งปากมดลูกจากการใช้เว็บแอปพลิเคชัน Cervical Cancer	4.35	มากที่สุด
เว็บแอปพลิเคชัน Cervical Cancer มีความเร็วในการตอบสนองต่อการใช้งานได้เป็นอย่างดี	4.10	มาก

มากที่สุด	ค่าเฉลี่ย 4.21 – 5.00
มาก	ค่าเฉลี่ย 3.41 – 4.20
ปานกลาง	ค่าเฉลี่ย 2.61 – 3.40
น้อย	ค่าเฉลี่ย 1.81 – 2.60
น้อยที่สุด	ค่าเฉลี่ย 1.00 – 1.80

### อภิปรายผล

ผลลัพธ์ของกฎที่ได้จากการทำเหมืองข้อมูลแบบ Classification ในการวิจัยครั้งนี้ ได้นำผลลัพธ์ที่ได้มาเขียนเพื่อใช้ในเว็บแอปพลิเคชัน เพื่อเป็นการวิจัยความเสี่ยงของโรคมะเร็งปากมดลูก ข้อมูลที่ได้มาจากโรงพยาบาล Universitario de Caracas เมืองการากัส ประเทศเวเนซุเอลา โดยใช้อัลกอริทึมที่มีความแม่นยำคือ REP ทำได้แค่วัดความเสี่ยงเบื้องต้นเท่านั้น นอกจากนี้ยังมีความรู้วิธีการป้องกันความเสี่ยงเบื้องต้นของโรคมะเร็งปากมดลูก ผ่านเว็บแอปพลิเคชันที่ผู้วิจัยได้สร้างเพื่อบอกถึงความเสี่ยงของผู้ใช้สามารถนำไปพัฒนาต่อให้ดีขึ้น และงานวิจัยฉบับนี้ได้สอดคล้องกับงานวิจัยของ กิตติศักดิ์ สุมามาลย์ ตามที่ได้กล่าวมาข้างต้น

### สรุปผล

ผลการวิจัยจากอัลกอริทึมที่มีค่ามากที่สุด คือ REP TREE กฎการจำแนกที่มีค่า Cross-validation เท่ากับ 10 folds ค่า Accuracy 0.989, Precision 0.988, Recall 1.00 และค่า F-Measure 0.994 โดยผลที่ได้จากการวิเคราะห์นั้น สอดคล้องกับวัตถุประสงค์ ดังนี้ 1) เพื่อวิเคราะห์ปัจจัยที่ส่งผลให้เกิดความเสี่ยงโรคมะเร็งปากมดลูก ซึ่งได้แก่ Schiller (การวินิจฉัยทางเคมี) Cytology (การวินิจฉัยช่องท้อง หรือ ปัสสาวะ) Dx: Cancer (การวินิจฉัยโรคมะเร็ง) และ Biopsy (การวินิจฉัยตรวจชิ้นเนื้อ) เป็นต้น 2) เพื่อพัฒนาระบบคัดกรองของผู้ที่เสี่ยงต่อการเกิดโรคมะเร็งปากมดลูก ซึ่งได้ออกแบบเว็บแอปพลิเคชันสำหรับการคัดกรองความเสี่ยง 3) เพื่อการศึกษาให้ความรู้เบื้องต้นของโรคมะเร็งปากมดลูก ข้อมูลจะอยู่ในเว็บแอปพลิเคชันมี 3 หัวข้อหลัก ได้แก่ ความรู้เบื้องต้นของโรคมะเร็งปากมดลูก วิธีการป้องกันโรคมะเร็งปากมดลูก และ แบบประเมินตรวจสอบความเสี่ยงโรคมะเร็งปากมดลูก

เอกสารอ้างอิง

- กิตติศักดิ์ สุมาลย์. (2555). *การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล*. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิศวกรรมเว็บ. กรุงเทพมหานคร: มหาวิทยาลัยธุรกิจบัณฑิต.
- จตุพล ศรีสมบุญ. (2555). *มะเร็งปากมดลูกป้องกันได้*. สืบค้นเมื่อ 15 กุมภาพันธ์ 2563, จาก <http://www.tgcsthai.com/dataview.php?id=240&code=KP>
- ณัฐวดี หงส์บุญมี และ พงศ์รินทร์ ศรีรุ่ง. (2561). *การประยุกต์ใช้เทคนิคจำแนกข้อมูลแบบต้นไม้ตัดสินใจเพื่อการวินิจฉัยโรคในโคเบื้องต้นบนโทรศัพท์มือถือ*. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี, 20(1), 7-14.
- ธนพัชร ดีทองอ่อน. (2552). *โรคมะเร็งปากมดลูกภัยร้ายใกล้ตัวผู้หญิง*. สืบค้นเมื่อ 15 กุมภาพันธ์ 2563, จาก <https://www.sikarin.com/content/detail/1>
- ไพฑูรย์ ออบเชย. (2553). *อุบัติการณ์มะเร็งปากมดลูกในประเทศไทย*. สืบค้นเมื่อ 20 พฤศจิกายน 2562, จาก [https://stri.cmu.ac.th/article\\_detail.php?id=17](https://stri.cmu.ac.th/article_detail.php?id=17)
- สายชล สิ้นสมบุญทอง. (2563). *การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล*. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์, 28(3), 4-9.
- สุดาฟ้า วงศ์หาริมาศย์. (2556). *ปัจจัยที่เกี่ยวข้องกับการไม่มาตรวจคัดกรองมะเร็งปากมดลูกในชุมชนที่คัดสรร*. วารสารวิชาการกรมสนับสนุน บริการสุขภาพ, 9(1), 12-20.
- Kelwin Fernandes, Jaime S. Cardoso, and Jessica Fernandes. (2017). *Cervical Cancer Data Set*. สืบค้นเมื่อ 18 พฤศจิกายน 2562 , จาก <https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer>