The 26th Annual Meeting in Mathematics
and
The 1st International Annual Meeting in Mathematics 2022
(AMM 2022)

Conference Proceedings

"Frontiers in Mathematics for Smart and Sustainable Development"

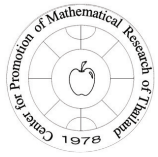
คณิตศาสตร์แนวหน้าสำหรับการพัฒนาแบบฉลาดและยั่งยืน

May 18-20, 2022

การจัดประชุมวิชาการนานาชาติ ครั้งที่ 1
จัดโดย สมาคมคณิตศาสตร์แห่งประเทศไทยในพระบรมราชูปถัมภ์
ร่วมกับ สาขาวิชาคณิตศาสตร์ สำนักวิชาวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี
และสาขาวิชาคณิตศาสตร์และสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครราชสีมา

AMM 2022

การประชุมวิชาการทางคณิตศาสตร์ ครั้งที่ 26 ประจำปี 2565



รายงานการประชุม

สารบัญ -- Table of Contents

รายนามผู้สนับสนุน -- List of Sponsors 1

สารจากอธิการบดี มหาวิทยาลัยเทคโนโลยีสุรนารี 2
Message from the Rector, Suranaree University of Technology

สารจากนายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ 4
Message from the President of the Mathematical Association of Thailand Under Patronage of His Majesty the King

สารจากอธิการ มหาวิทยาลัยราชภัฏนครราชสีมา 5
Message from the Rector of Nakhon Ratchasima Rajabhat University

สารจากผู้อำนวยการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย 6
Message from the Director of the Center for Promotion of Mathematical Research of Thailand

สารนำจากคณะกรรมการจัดการประชุมฯ
Message from the Conference Organizing Committee 7

สารบัญ -- Table of Contents 8

กำหนดการ -- Conference Schedule 9

Keynote Speakers 14

Invited Speakers 17

Part I: Proceedings of the 26th Annual Meeting in Mathematics 20

Part II: Proceedings of the 1st International Annual Meeting in Mathematics 417

คณะกรรมการจัดการประชุมฯ --- Conference Committees A

Proceedings of the 26th Annual Meeting in Mathematics

สารบัญ -- Table of Contents

01 Algebra and Number Theory (AN)

AN-N-01 Solutions to a Quadratic Equation over Finite Fields 1
Puchong Wongkumpra and Detchat Samart

AN-N-08 The Weak First Exponential Law of Mixed Product of *n*-ary Hyperalgebras Carried by Good Homomorphisms 12
Nitima Phrommarat and Thanwarat Butsan

AN-N-10 Iterative Algorithm for Polynomial Modular Inversion Modulo $x^{p^r} - 1$ Over Finite Field of order *p* 21
Samakorn Sripatthanakul and Wutichai Chongchitmate

02 Geometry and Graph Theory (GG)

GG-N-01 Some Families of Self-Clique Graphs whose Clique Size Sequence is (2,...,2,3,3,3) 27
Sajika Tubtim, Sirirat Singhun and Ratinan Boonklurb

GG-N-02 Charges of Semistandard Young Tableaux of Certain Shapes and Contents 36
Nattanon Tualue and Ouamporn Phuksuwan

04 Differential Equations and Dynamical Systems (DE)

DE-N-01 Algebraic Independence of Solutions of Certain Second Order Homogeneous Linear Differential Equations with Linear Coefficients 50
Phisitphong Ketrat and Vichian Laohakosol

DE-N-03 Beyond Quenching Profile for Singular Semilinear Parabolic Problem With Mixed Boundary Conditions 62
Benjamin Thaitavorn and Ratinan Boonklurb

05 Mathematical Modeling and Numerical Mathematics (MN)

MN-N-01 Numerical Simulation of Water Quality in a Couple of Ponds of Shrimp Farming 70
Nattinee Sittijinda and Nopparat Pochai

MN-N-02	A Non-Dimensional Mathematical Model of Shoreline Evolution with a Groin Structure *Surasak Manilam and Nopparat Pochai*	81
MN-N-03	การเลือกตำแหน่งที่ตั้งและจำนวนหัวชาร์จของรถยนต์ไฟฟ้าที่เหมาะสมในกรุงเทพมหานคร *ณิชากร ชัยบัญชากิจ, พศิกาญจน์ ทับประดง, ณัฏฐ์ เหลืองสวัสดิพร, และ สายฝน จาตุรันตบุตร*	90
MN-N-05	แบบจำลองการแพร่กระจายผู้ติดเชื้อโควิด-19 ภายใต้ภาวะควบคุมการระบาดด้วยการฉีดวัคซีน *ธนาทร อินทรปัญญา, อภิชาต ศุรธณี, สิทธิพร ภัทรดิลกรัตน และ กิติพร พลายมาศ*	115
MN-N-06	A Mathematical Model for Measuring Carbon Dioxide Concentration in a Bus Due to Passengers Breathing *Jenjira Sooknum and Nopparat Pochai*	132
MN-N-07	Numerical Algorithm Based on Finite Integration Method Using Shifted Chebyshev Expansion for Solving Moving Boundary Problems *Warunya Wong-u-ra and Ratinan Boonklurb*	141
MN-N-08	The Simplex Method with the Minimal Angle Jump for Solving Linear Programming Problems *Monsicha Tipawanna and Krung Sinapiromsaran*	153
MN-N-09	โจทย์ปัญหาทางคณิตศาสตร์ที่น่าสนใจและ การประยุกต์ใช้โปรแกรมภาษา C++ เพื่อช่วยใน การหาผลเฉลยของปัญหาทางคณิตศาสตร์ *อมรรัตน์ สุริยวิจิตรเศรณี และ เจษฎา ตัณฑนุช*	163

06 Probability Theory (PR)

PR-N-01	การเปรียบเทียบการทดสอบค่ากลางของประชากรสองกลุ่มที่เป็นอิสระกันเมื่อตัวอย่างทีขนาดเล็กมาก *มนต์นภา พงษ์พรรณากูล และ วนิดา พงษ์ศักดิ์ชาติ*	176
PR-N-02	An Improvement of the Error Bound of Local Limit Theorems for Sums of Independent Lattice Random Variables *Punyapat Kammoo, Kritsana Neammanee and Kittipong Laipaporn*	187

PR-N-03	Poisson Approximation for Sums of Independent Non-Negative Integer-valued Random Variables *Supavit Kiatteerarat, Kritsana Neammanee and Suporn Jongpreechaharn*	197
PR-N-04	การเปรียบเทียบประสิทธิภาพของแผนภาพกล่องสำหรับการตรวจสอบค่านอกเกณฑ์ *ทศวรรษ ณ บางชาง และ บำรุงศักดิ์ เผื่อนอารีย*	206
PR-N-05	A Local Limit Theorem for Negative Binomial Random Sums *Hattacha Kongjiw, Petcharat Rattanawong and Kritsana Neammanee*	218

07 Data Science and Statistics (DS)

DS-N-01	โครงข่ายประสาทเทียมสำหรับการวิเคราะห์การถดถอยเชิงเส้นตามบริบทนัยทั่วไป *ชยานนท์ ขัตติยาภิรักษ์ และ เสกสรร เกียรติสุไพบูลย*	228
DS-N-02	การทำนายราคาของหลักทรัพย์โดยวิศวกรรมคุณลักษณะและเทคนิคการเรียนรู้ของเครื่อง *รัชพล ปริโยทัย และ เบญจวรรณ โรจนดิษฐ*	243
DS-N-04	โครงข่ายประสาทเทียมเชิงพยากรณ์แบบปรับปรุงโดยใช้การเลือกสับเซตที่ดีที่สุด *พรชิตา ธนากร, สุปราณี ลิสวัสดิ์ และ พัทธ์ชนก ศรีสุรเดชชัย*	263
DS-N-05	การทำนายโรคเบาหวานโดยใช้วิศวกรรมคุณลักษณะสำหรับขั้นตอนวิธีการจำแนกในการเรียนรู้ของเครื่อง *คุณาภรณ์ พันธ์เพียร, จักรกฤษณ์ พลรบ และ เจษฎา ตัณฑนุช*	278
DS-N-06	การเฉลี่ยตัวแบบบนต้นไม้การถดถอยสำหรับการพยากรณ์ *ชนินทร แกววิบูลยพันธุ์, สุปราณี ลิสวัสดิ์ และ พัทธชนก ศรีสุรเดชชัย*	290
DS-N-08	การวิเคราะห์พื้นที่ที่เหมาะสมสำหรับการปลูกยางพาราตามปัจจัยสภาพภูมิอากาศด้วยวิธีการตัดสินใจแบบหลายหลักเกณฑ์ *พรทิพย์ เดชพิชัย, กัญญาวีร์ คำสนอนันตกุล, ธัญชนก ชัยกูล และ บิลกีส วงษ์พิรำ*	316
DS-N-09	การวิเคราะห์ความรู้สึกที่มีต่อการทำประกันภัยด้วยเทคนิคการเรียนรู้เชิงลึก กรณีศึกษากระทู้ออนไลน์พันทิป *คงภพ ไชยศร และ บุษยมาส พิมพ์พรรณชาติ*	321

DS-N-10	การเปรียบเทียบตัวแบบ SARIMAX และตัวแบบแยกส่วนประกอบร่วมกับ SARIMAX ในการพยากรณ์ค่าสินไหมทดแทนของธุรกิจประกันภัยรถยนต์ในประเทศไทย *นิฉา แกวหาวงษ์, ไอริณ สิมสีแก้ว, ณัฐกิตติ์ การเร็ว และ วรรณกานต์ วงศ์เสนา*	332
DS-N-11	การศึกษาและพัฒนาระบบการจัดการงานสินไหมในกรณีความเสียหายหนัก ของการประกันภัยรถยนต์ *เบญจมาภรณ์ ศรีอัมพร, บุษยมาส พิมพ์พรรณชาติ และ เทิดขวัญ ช้างเผือก*	346

08 Mathematics Education (ED)

ED-N-01	การพัฒนาความเข้าใจทางคณิตศาสตร์ของนักเรียนโดยใช้การศึกษาชั้นเรียนและวิธีการแบบเปิด *กนกวรรณ รัตนจำนอง, วิภาพร สุทธิอัมพร และ สุริพร บุญเมือง*	358
ED-N-02	การพัฒนาชุดการเรียนรู้ด้วยตนเองเรื่องเรขาคณิตสำหรับผู้เรียนในระดับมัธยมศึกษาตอนปลายโดยเปรียบเทียบมโนทัศน์เรขาคณิตแบบฉบับและเรขาคณิตวิเคราะห์ *วุฒิชัย ไชยปญญา และ ศุภณัฐ ชัยดี*	366
ED-N-03	การคิดทางคณิตศาสตร์ของนักเรียนในชั้นเรียนที่จัดการเรียนรู้แบบผสมผสานด้วยวิธีการแบบเปิด *ธัญญารัตน์ ถ่องแท้, สุดาทิพย์ หาญเชิงชัย และศศลักษณ์ ขลิกดำ*	381

Proceedings of the
26th Annual Meeting in Mathematics (AMM 2022)
School of Mathematics, Institute of Science
Suranaree University of Technology, Thailand

การทำนายโรคเบาหวานโดยใช้วิศวกรรมคุณลักษณะสำหรับขั้นตอนวิธีการจำแนกในการเรียนรู้ของเครื่อง

คุณาภรณ์ พันธ์เพียร[1,†] จักรกฤษณ์ พลรบ[2] และ เจษฎา ตัณฑนุช[2,‡]

[1]สาขาวิชานวัตกรรม วิศวชีวการแพทย์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี 30000

[2]สาขาวิชาคณิตศาสตร์ สำนักวิชาวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี 30000

บทคัดย่อ

จุดมุ่งหมายของการศึกษานี้คือ เพื่อศึกษาปัจจัยเสี่ยงที่มีผลกระทบต่อการทำให้เกิดโรคเบาหวาน โดยใช้วิธีวิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม แล้วนำไปพัฒนาแบบจำลองของผู้ป่วยโรคเบาหวาน ซึ่งการศึกษาครั้งนี้ได้นำข้อมูลจากการตอบแบบสำรวจจำนวน 70,692 รายการ โดยได้ข้อมูลจากเว็บไซต์ Kaggle แล้วนำมาใช้เทคนิควิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่มเพื่อหาคุณลักษณะที่ดีที่สุดออกมา จากนั้นนำไปสร้างแบบจำลองด้วย 5 ขั้นตอนวิธี ได้แก่ ซัพพอร์ทเวกเตอร์แมชชีน, ต้นไม้ตัดสินใจ, เทคนิคเพื่อนบ้านใกล้ที่สุด, เกรเดียนท์บูตทรี และป่าสุ่ม แล้วทำการเปรียบเทียบกับการทำแบบจำลองด้วย 5 ขั้นตอนวิธีดังกล่าวกับข้อมูลต้นฉบับโดยตรง ผลการศึกษาพบว่า คุณลักษณะในข้อมูลต้นฉบับมีทั้งหมด 21 คุณลักษณะ หลังจากใช้การสกัดจากวิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม ทำให้เหลือคุณลักษณะที่ใช้พิจารณา 7 คุณลักษณะ ได้แก่ 1) ความดันโลหิตสูง 2) สุขภาพโดยทั่วไป 3) ค่าดัชนีมวลกาย 4) อายุ 5) เงินเดือน 6) คลอเรสเตอรอลสูง และ 7) เพศ และการใช้เทคนิควิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม แล้วสร้างแบบจำลองด้วยขั้นตอนวิธีซัพพอร์ทเวกเตอร์แมชชีน, เทคนิคเพื่อนบ้านใกล้ที่สุด และเกรเดียนท์บูตทรีให้ค่าความแม่นยำดีขึ้น แต่อีก 2 ขั้นตอนวิธี ให้ค่าความแม่นยำที่ลดลงกว่าการทำแบบจำลองกับข้อมูลต้นฉบับโดยตรงแต่ยังคงมีค่าใกล้เคียงกัน ในการสร้างแบบจำลองทั้ง 10 แบบพบว่า การสร้างแบบจำลองด้วยวิธีป่าสุ่ม กับข้อมูลต้นฉบับให้ค่าความแม่นยำสูงที่สุดคือ 74.07% และการสร้างแบบจำลองด้วยวิธีป่าสุ่ม กับข้อมูลที่ผ่านการใช้เทคนิควิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่ม ให้ค่าความแม่นยำรองลงมาคือ 73.68%

คำสำคัญ: โรคเบาหวาน, วิศวกรรมคุณลักษณะ, การเรียนรู้ของเครื่อง

2020 MSC: ปฐมภูมิ 68T99 ทุติยภูมิ 92C50

[†]ผู้นำเสนอ [‡]ผู้แต่งหลัก

อีเมล: kunaporn.m6300609@gmail.com, jessada@g.sut.ac.th, m6300289@g.sut.ac.th

1 บทนำ

โรคเบาหวานเป็นปัญหาทางสาธารณสุขที่สำคัญโรคหนึ่งของทั่วโลกรวมทั้งประเทศไทยด้วย โดยเป็นโรคที่ร่างกายมีระดับน้ำตาลในเลือดสูงกว่าปกติ นอกจากนี้ยังคงเป็นสาเหตุหลักที่ก่อให้เกิดภาวะแทรกซ้อนต่าง ๆ เช่น โรคแทรกซ้อนทางตา โรคแทรกซ้อนทางไต จากสถานการณ์โรคเบาหวานในปี 2564 พบว่าทั่วโลกมีผู้ป่วยจำนวน 463 ล้านคน และคาดการณ์ว่าในปี 2588 จะมีผู้ป่วยโรคเบาหวานเพิ่มขึ้นถึง 629 ล้านคน สำหรับประเทศไทยพบอุบัติการณ์โรคเบาหวานมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง มีผู้ป่วยรายใหม่เพิ่มขึ้นประมาณ 3 แสนคนต่อปี และมีผู้ป่วยโรคเบาหวานอยู่ในระบบทะเบียนของกระทรวงสาธารณสุขจำนวน 3.2 ล้านคน ก่อให้เกิดการสูญเสียค่าใช้จ่ายในการรักษาด้านสาธารณสุขอย่างมหาศาล เฉพาะเบาหวานเพียงโรคเดียวทำให้สูญเสียค่าใช้จ่ายในการรักษาเฉลี่ยสูงถึง 47,596 ล้านบาทต่อปี และหากรวมอีก 3 โรค คือ โรคความดันโลหิตสูง โรคหัวใจ และโรคหลอดเลือดสมอง ทำให้ภาครัฐสูญเสียงบประมาณในการรักษารวมกันสูงถึง 302,367 ล้านบาทต่อปี โดยโรคเบาหวานก่อให้เกิดโรคแทรกซ้อนมากมาย และมีอัตราการเสียชีวิตเนื่องจากโรคเบาหวานเฉลี่ยวันละ 200 คน [1]

การนำเทคโนโลยีมาประยุกต์ใช้ในการวินิจฉัยโรคต่าง ๆ รวมไปถึงการใช้เทคโนโลยีทางด้านการเรียนรู้ของเครื่อง (Machine Learning: ML) และ ปัญญาประดิษฐ์ (Artificial Intelligence: AI) มาช่วยในการวินิจฉัยโรคต่าง ๆ เพื่อพัฒนาให้การรักษามีประสิทธิภาพมากขึ้น เนื่องจากเป็นเทคโนโลยีที่มีความทันสมัยในปัจจุบัน และมีการมุ่งเน้นการใช้ตัวอย่างหรือประสบการณ์เพื่อการเรียนรู้งาน ทำให้มีระดับความแม่นยำในการวินิจฉัยโรคที่สูง ซึ่งจะเป็นกำลังสำคัญเข้ามาช่วยสนับสนุนการทำงานของแพทย์ ดังนั้นผลลัพธ์ที่ได้จากการทำงานร่วมกันคือเพิ่มทั้งความเร็วในกระบวนการคัดกรองผู้ป่วย การตรวจสอบรักษา โดยผู้ป่วยก็จะไม่เสียโอกาสได้เข้ารับการรักษาตั้งแต่ช่วงระยะแรกที่เริ่มมีแนวโน้มเสี่ยงต่อการเกิดโรค นอกจากนี้หากได้เข้ารับการรักษาที่รวดเร็วยังช่วยลดการเกิดภาวะแทรกซ้อน และลดการสูญเสียอวัยวะต่าง ๆ ได้อีกด้วย

2 ความรู้พื้นฐาน

2.1 ความรู้เกี่ยวกับโรคเบาหวาน

โรคเบาหวานเป็นโรคที่มีความผิดปกติเกี่ยวกับการนำน้ำตาลไปใช้ประโยชน์อันเกี่ยวเนื่องกับความบกพร่องของฮอร์โมนอินซูลิน หรือประสิทธิภาพการทำงานของฮอร์โมนอินซูลินลดลง ส่งผลให้กระบวนการดูดซึมน้ำตาลในเลือดให้เป็นพลังงานของเซลล์ในร่างกายทำงานได้ไม่เต็มประสิทธิภาพ ทำให้ร่างกายมีระดับน้ำตาลในเลือดสูงกว่าปกติ หากน้ำตาลในกระแสเลือดสูงมากขึ้นถึงระดับหนึ่ง จะทำให้ไตซึ่งปกติจะมีหน้าที่ดูดกลับน้ำตาลจากสารที่ถูกกรองจากหน่วยไตไปใช้ ดูดกลับน้ำตาลได้ไม่หมด ส่งผลให้มีน้ำตาลรั่วออกมากับปัสสาวะ [2]

โรคเบาหวานในระยะแรกจะไม่แสดงอาการผิดปกติบางรายอาจตรวจพบโรคเบาหวานเมื่อพบภาวะแทรกซ้อนขึ้นแล้ว ซึ่งอาการที่พบส่วนใหญ่ คือ กระหายน้ำมาก ปากแห้ง ปัสสาวะบ่อย หิวบ่อย น้ำหนักลดหรือเพิ่มผิดปกติ สายตาพร่ามัว รู้สึกเหนื่อยง่าย ชาบริเวณปลายมือปลายเท้า บาดแผลหายยาก หย่อนสมรรถภาพทางเพศ [3]

ภาวะแทรกซ้อนหลัก ๆ ที่พบได้ในผู้ป่วยเบาหวาน มีดังนี้ [4]

- **ภาวะแทรกซ้อนทางตา** หรือที่เรียกกันว่า ภาวะเบาหวานขึ้นตา (Retinopathy) หรือ จอประสาทตาเสื่อม

- **ภาวะแทรกซ้อนทางไต** หรือที่เรียกกันว่า ภาวะเบาหวานลงไต ส่งผลให้เกิดโรคไตเสื่อม (Nephropathy) หรือไตวายเรื้อรัง (Chronic renal failure)
- **ภาวะแทรกซ้อนทางเส้นประสาท** มีอาการระบบประสาทเสื่อม (Neuropathy)
- **เส้นเลือดแดงใหญ่อุดตัน** อาการที่พบได้บ่อยคือ มีอาการปวดขามากเมื่อเดินหรือวิ่ง หากไม่ได้รับการรักษาเป็นเวลานานอาจทำให้เกิดภาวะเส้นเลือดอุดตันจนปลายเท้าขาดเลือด ติดเชื้อ และอาจต้องตัดนิ้วเท้าหรือขาทิ้งได้
- **เส้นเลือดหัวใจตีบ** เป็นภาวะแทรกซ้อนที่พบได้บ่อยในผู้ป่วยโรคเบาหวาน และเป็นภาวะแทรกซ้อนที่รุนแรง มีอาการหัวใจวาย ความดันโลหิตต่ำ หัวใจเต้นผิดจังหวะ และเสียชีวิตอย่างเฉียบพลันได้
- **เส้นเลือดสมองตีบ** เป็นภาวะแทรกซ้อนที่รุนแรงเช่นกัน เมื่อเกิดภาวะเส้นเลือดสมองตีบ ทำให้การทำงานของสมองและเส้นประสาทบริเวณที่ขาดเลือดลดลงหรือไม่ทำงาน ส่งผลให้เกิดอัมพฤกษ์ อัมพาต

ปัจจุบันการวินิจฉัยโรคเบาหวานในประเทศไทย แพทย์สามารถวินิจฉัยอาการเบื้องต้นของโรคเบาหวานได้จากประวัติอาการ ประวัติการเจ็บป่วยต่าง ๆ ประวัติการเป็นโรคเบาหวานของคนในครอบครัว และที่สำคัญมากที่สุดคือการตรวจเลือดเพื่อดูปริมาณน้ำตาลในเลือด (ผู้ป่วยโรคเบาหวานจะพบปริมาณน้ำตาลในเลือดมากกว่าคนปกติ) และมีการตรวจอื่น ๆ ประกอบไปด้วยตามความเหมาะสม เช่น การตรวจปัสสาวะเพื่อดูน้ำตาลในปัสสาวะ (ซึ่งจะไม่พบในคนปกติ) การตรวจเลือดเพื่อดูการทำงานของไต (เพราะโรคเบาหวานมักส่งผลต่อการเกิดแทรกซ้อนเช่น โรคไตเรื้อรัง) และการตรวจสุขภาพตาโดยจักษุแพทย์ (เพื่อเฝ้าระวังภาวะแทรกซ้อนของโรคเบาหวานต่อจอตา)

2.2 การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่องแบ่งออกเป็น 3 ประเภทดังนี้ [5]

1. การเรียนรู้แบบมีการสอน (Supervised Learning) การเรียนรู้ชนิดนี้ แบ่งออกเป็น 2 ประเภท คือการจำแนกประเภท (Classification) และการถดถอย (Regression)

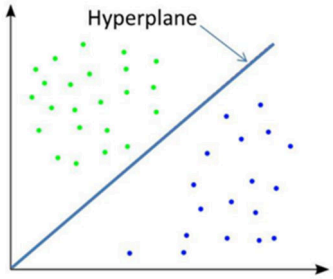
2. การเรียนรู้แบบไม่ต้องมีการสอน (Unsupervised Learning) การเรียนรู้ชนิดนี้ แบ่งออกเป็น 2 ประเภท คือการจัดกลุ่ม (Clustering) การหาความสัมพันธ์ (Association)

3. การเรียนรู้อาศัยการป้อนกลับผลลัพธ์ (Reinforcement Learning) ซึ่งใช้ผลจากการป้อนกลับมาปรับปรุงการเรียนรู้ของตัวเอง

2.3 แบบจำลองสำหรับการจำแนกประเภท (Models for Classification)

2.3.1 ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine (SVM))

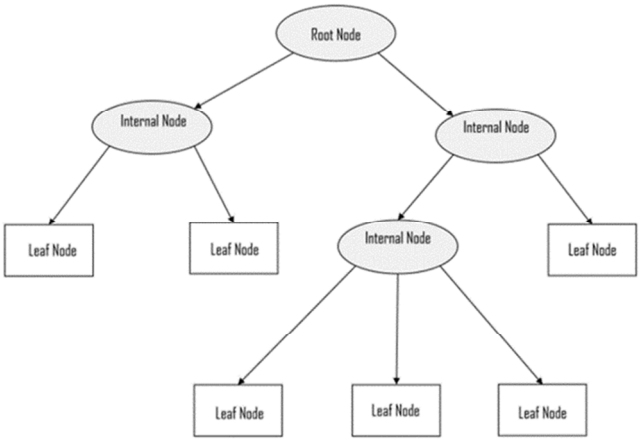
หลักการทำงานคือการนำค่าของกลุ่มข้อมูลมาพล็อตลงในพื้นที่เรียกว่าฟีเจอร์สเปซ (Feature Space) จากนั้นหาแนวขอบเขตการตัดสินใจที่ดีที่สุดเรียกว่าไฮเปอร์เพลน (Hyper plan) ที่ใช้แบ่งข้อมูลกลุ่มออกจากกัน โดยจะสร้างแนวที่ดีที่สุดในการแบ่งกลุ่ม



รูปที่ 1 ซัพพอร์ทเวกเตอร์แมชชีน [6]

2.3.2 ต้นไม้ตัดสินใจ (Decision Tree)

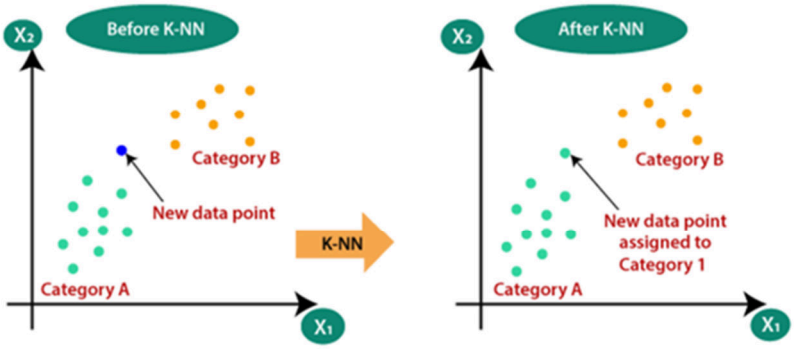
Decision Tree หรือที่เรียกว่าต้นไม้ตัดสินใจ หลักการทำงานคล้ายกับการถามตอบ เป็นการเรียนรู้จากคุณลักษณะของข้อมูล (Attributes) แล้วสร้างผังการตัดสินใจคล้ายกับต้นไม้ จึงเรียกว่าต้นไม้ตัดสินใจ โดยผลลัพธ์จะมีเพียงสองกลุ่ม หรือมากกว่าก็ได้



รูปที่ 2 โครงสร้างของต้นไม้ตัดสินใจ [7]

2.3.3 เทคนิคเพื่อนบ้านใกล้ที่สุด (k – Nearest Neighbors (kNN))

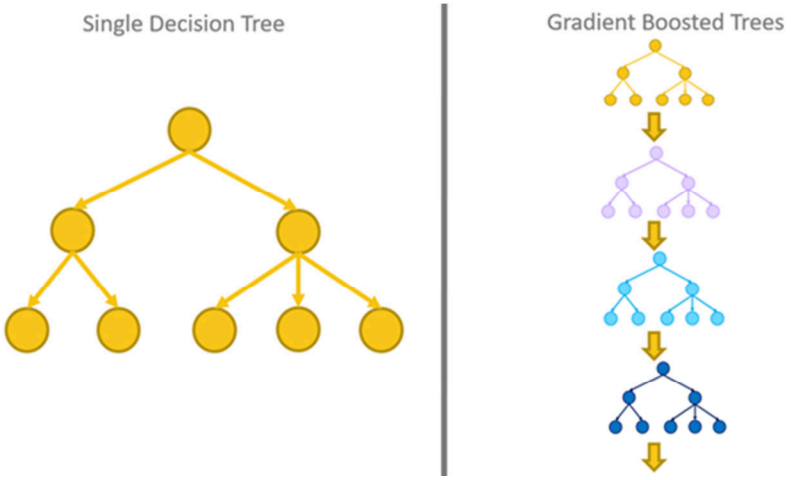
หลักการทำงานคือการเทียบหาข้อมูลจุดใหม่ ถ้าพบว่าอยู่ใกล้กับกลุ่มใด ก็จะจัดให้ข้อมูลใหม่อยู่ในกลุ่มนั้น หมายความว่าเมื่อมีข้อมูลใหม่ปรากฏขึ้น ก็สามารถจำแนกประเภทข้อมูลได้ง่ายโดยใช้ขั้นตอนวิธี k-NN (ชื่อ Nearest Neighbors หมายถึงจัดให้เข้ากลุ่มกับเพื่อนบ้านที่ใกล้ที่สุด)



รูปที่ 3 เทคนิคเพื่อนบ้านใกล้ที่สุด [8]

2.3.4 เกรเดียนท์บูตทรี (Gradient Boosted Trees)

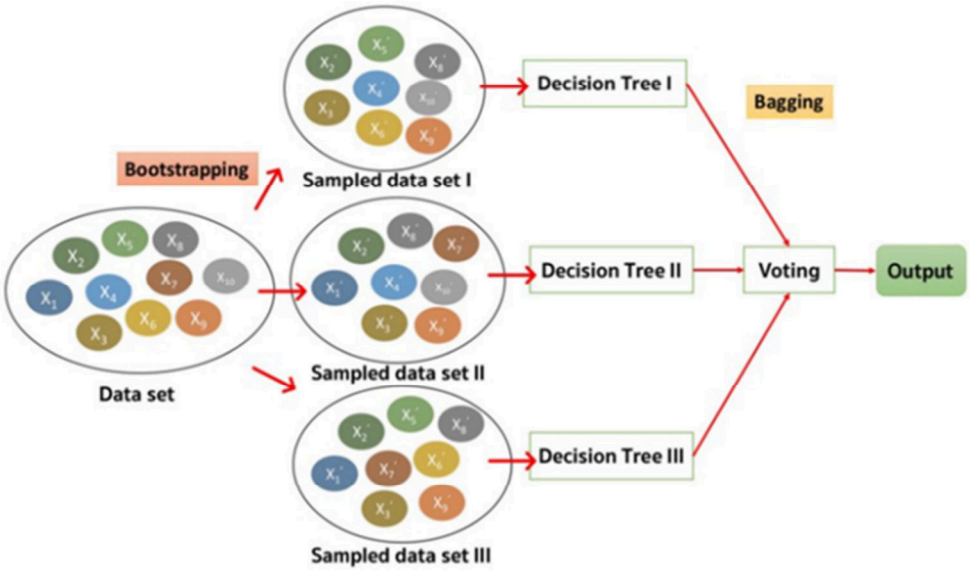
เป็นขั้นตอนวิธีที่มีพื้นฐานมาจากต้นไม้ตัดสินใจซึ่งเป็นการปรับปรุงประสิทธิภาพให้มีค่าสูงขึ้น โดยการสุ่มสร้างต้นไม้ตัดสินใจหลายร้อยแบบจำลอง และประเมินผลแต่ละแบบจำลองจนกว่าจะได้ต้นไม้ตัดสินใจที่สมบูรณ์



รูปที่ 4 เกรเดียนท์บูตทรี [9]

2.3.5 ป่าสุ่ม (Random Forest)

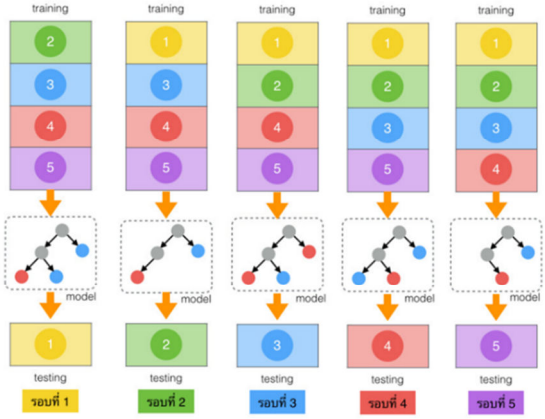
ป่าสุ่มถูกพัฒนาขึ้นจากต้นไม้ตัดสินใจ ต่างกันที่ป่าสุ่มเป็นการเพิ่มจำนวน ต้นไม้เป็นหลาย ๆ ต้น ทำให้มีประสิทธิภาพในการทำงานสูงขึ้น แม่นยำมากขึ้น หลักการป่าสุ่มคล้ายต้นไม้หรือต้นไม้ตัดสินใจปกติ แต่จะสุ่มเอาข้อมูล (Instance) ไปสร้างเป็นต้นไม้หลาย ๆ ต้น แต่ละต้นเรียกว่า Subset เหมือนกับว่ามีป่าที่มีต้นไม้จำนวนมาก ๆ ซึ่งแต่ละต้นจะมีรูปแบบสุ่มไม่เหมือนกัน ในตอนทำงานจะให้แต่ละต้นทำนาย และคำนวณผลการทำนายด้วยการโหวตผลลัพธ์ที่ถูกเลือกมากที่สุด



รูปที่ 5 โครงสร้างของป่าสุ่ม [10]

2.4 การทดสอบแบบไขว้ (K-fold Cross validation)

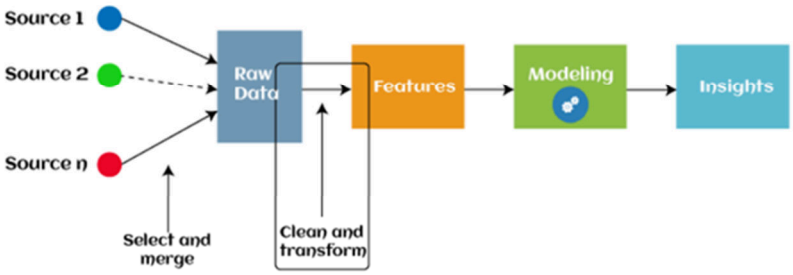
วิธีการทดสอบการทดสอบแบบไขว้ เป็นวิธีที่นิยมในการทำงานวิจัย เพื่อใช้ในการทดสอบประสิทธิภาพของแบบจำลองเนื่องจากผลที่ได้มีความน่าเชื่อถือ ลักษณะการทำงานคือแบ่งข้อมูลออกเป็นหลายส่วน มักจะแสดงด้วยค่า K



รูปที่ 6 การทดสอบแบบจำลองด้วยวิธี 5-การทดสอบแบบไขว้ [11]

2.5 วิศวกรรมคุณลักษณะ (Feature Engineering)

วิศวกรรมคุณลักษณะเป็นขั้นตอนก่อนการประมวลผลของการเรียนรู้ของเครื่อง ซึ่งจะดึงคุณลักษณะจากข้อมูลดิบ ช่วยแสดงปัญหาพื้นฐานของแบบจำลองการคาดการณ์ในทางที่ดีขึ้น ซึ่งจะส่งผลให้สามารถปรับปรุงความถูกต้องของแบบจำลองสำหรับข้อมูลที่ไม่เคยเห็นมาก่อน (unseen data) แบบจำลองการทำนายประกอบด้วยตัวแปรทำนายและตัวแปรผลลัพธ์ และในขณะที่กระบวนการวิศวกรรมคุณลักษณะจะเลือกตัวแปรการทำนายที่มีประโยชน์ที่สุดสำหรับแบบจำลอง [12]



รูปที่ 7 กระบวนการทำงานของวิศวกรรมคุณลักษณะ [12]

วิศวกรรมคุณลักษณะในการเรียนรู้ของเครื่อง ประกอบด้วย 6 กระบวนการหลัก ดังนี้

1. การเตรียมข้อมูล (Data Preparation)
2. การสร้างคุณลักษณะ (Feature Generation)
3. การแปลงคุณลักษณะ (Feature Transformation)
4. การแยกคุณลักษณะ (Feature Extraction)
5. การเลือกคุณลักษณะ (Feature Selection)
6. วิศวกรรมคุณลักษณะอัตโนมัติ (Automatic Feature Engineering)

2.6 ตัวชี้วัดประสิทธิภาพ (Performance Metrics)

Accuracy คือ ความแม่น, Precision คือ ความเที่ยง, Specificity คือ อัตราผลลบจริง (True Negative Rate), Sensitivity คือ ค่าเรียกคืน (recall) หรือ อัตราผลบวกจริง (True Positive Rate), ROC AUC คือ ค่าพื้นที่ใต้เส้นโค้ง ROC (Area under the ROC curve), Kappa คือ ค่าสัมประสิทธิ์ Cohen (Cohen's kappa Coefficient) และ F1 score คำนวณโดย

$$Accuracy = \frac{True\ Positive\ (TP)\ +\ True\ Negative\ (TN)}{All} \tag{1}$$

$$Precision\ = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)\ +\ False\ Positive\ (FP)} \tag{2}$$

$$Specificity\ (TNR)\ = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN)\ +\ False\ Positive\ (FP)} \tag{3}$$

$$Sensitivity\ (TPR) = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)\ +\ False\ Negative\ (FN)} \tag{4}$$

$$FPR\ = \frac{False\ Positive\ (FP)}{True\ Negative\ (TN)\ +\ False\ Positive\ (FP)} \tag{5}$$

$$F1\ score\ =\ \frac{2 \times (Precision \times Recall)}{(Precision\ +\ Recall)} \tag{6}$$

$$\kappa = \frac{p_a - p_e}{1 - p_e} = 1 - \frac{1 - p_a}{1 - p_e} \tag{7}$$

หมายเหตุ เส้นโค้ง ROC เกิดจากการพล็อตกราฟระหว่าง แกน x คือ อัตราผลบวกเท็จ (False Positive Rate) และ แกน y คือ อัตราผลบวกจริง (True Positive Rate)

3 ขั้นตอนวิธีวิจัย

1. นำชุดข้อมูลจาก https://www.kaggle.com/alexteboul/diabetes-health-indicators-dataset ซึ่งมีจำนวนข้อมูลทั้งหมด 70,692 รายการ มีคุณลักษณะ 21 คุณลักษณะ และมีตัวแปรตาม คือผู้ป่วยที่เป็นโรคเบาหวานและไม่เป็นโรคเบาหวาน ซึ่งแสดงดังตารางที่ 1
2. แบ่งชุดข้อมูลออกเป็น 2 ส่วน โดยส่วนที่หนึ่ง คือ ชุดฝึกสอนและส่วนที่สอง คือ ชุดทดสอบ ซึ่งจะแบ่งในอัตราส่วน 75/25 (ชุดฝึกสอน/ชุดทดสอบ)
3. นำชุดฝึกสอนไปผ่านกระบวนการสกัดคุณลักษณะที่สำคัญ ด้วยวิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม
4. นำชุดฝึกสอนที่ได้สกัดคุณลักษณะเรียบร้อยแล้ว นำไปสร้างแบบจำลอง 5 แบบจำลอง ดังนี้ ซัพพอร์ทเวกเตอร์แมชชีน, ต้นไม้ตัดสินใจ, เทคนิคเพื่อนบ้านใกล้ที่สุด, เกรเดียนท์บูตทรี และป่าสุ่ม โดยในแต่ละแบบจำลองจะมีการหาพารามิเตอร์ที่ดีที่สุดของแบบจำลองนั้น ๆ ด้วยการประเมินแบบ 5-5-การทดสอบแบบไขว้ นำแบบจำลองที่ได้จากขั้นตอนที่ 4 มาทดสอบประสิทธิภาพกับชุดทดสอบ และวัดผลด้วยตัวชี้วัดประสิทธิภาพต่าง ๆ ดังนี้ Accuracy, Precision, Specificity, Sensitivity, F1 Score, ROC AUC Score และ Cohen's kappa Coefficient
5. เปรียบเทียบผลลัพธ์ระหว่างการใช้วิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่มแล้วสร้างแบบจำลองกับการใช้ชุดข้อมูลฝึกสอน โดยไม่มีการสกัดคุณลักษณะ แล้วสร้างแบบจำลอง

4 ผลการศึกษา

ผลการศึกษาที่ได้มีดังนี้

ตารางที่ 1 คุณลักษณะที่เกี่ยวข้องในชุดข้อมูลต้นฉบับ

1. ความดันโลหิตสูง	8. การออกกำลังกาย	15. สุขภาพจิต
2. คลอเรสเตอรอลสูง	9. การรับประทานผลไม้	16. สุขภาพทางกายภาพ
3. การตรวจคลอเรสเตอรอล	10. การรับประทานผัก	17. การเดินลำบาก
4. ค่าดัชนีมวลกาย	11. การดื่มแอลกอฮอล์	18. เพศ
5. สูบบุหรี่	12. ประกันสุขภาพ	19. อายุ
6. โรคหลอดเลือดสมอง	13. ไม่ไปพบแพทย์เพราะค่าใช้จ่าย	20. การศึกษา
7. โรคหลอดเลือดหัวใจ	14. สุขภาพโดยทั่วไป	21. เงินเดือน

จากการนำวิศวกรรมคุณลักษณะทำงานร่วมกับวิธีป่าสุ่มทำให้คัดเลือกเหลือเพียง 7 คุณลักษณะ ดังแสดงในตารางที่ 2

ตารางที่ 2 คุณสมบัติที่ได้จากการคัดเลือกโดยขั้นตอนวิธีวิศวกรรมคุณลักษณะทำงานร่วมกับวิธีป่าสุ่ม

1. ความดันโลหิตสูง	5. เงินเดือน
2. สุขภาพโดยทั่วไป	6. คลอเรสเตอรอลสูง
3. ค่าดัชนีมวลกาย	7. เพศ
4. อายุ	

ประสิทธิภาพจากการสร้างแบบจำลองด้วยขั้นตอนวิธีต่าง ๆ ได้แสดงในตารางที่ 3-7 ดังต่อไปนี้

ตารางที่ 3 ประสิทธิภาพการทำงานของซัพพอร์ทเวกเตอร์แมชชีนโดยเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ผ่านการสกัดจากวิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่ม

ตัวชี้วัดประสิทธิภาพ	**ชุดข้อมูลต้นฉบับ**	**วิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม**
Accuracy	68.52%	73.45%
Precision	61.94%	68.53%
Specificity	40.93%	60.17%
Sensitivity	96.11%	86.72%
F1 Score	75.33%	76.56%
ROC AUC Score	0.818	0.817
Kappa	0.370	0.469

ตารางที่ 4 ประสิทธิภาพการทำงานของต้นไม้ตัดสินใจโดยเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ผ่านการสกัดจากวิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่ม

ตัวชี้วัดประสิทธิภาพ	**ชุดข้อมูลต้นฉบับ**	**วิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม**
Accuracy	73.37%	73.10%
Precision	71.61%	71.64%
Specificity	69.31%	69.71%
Sensitivity	77.43%	76.49%
F1 Score	74.41%	73.99%
ROC AUC Score	0.806	0.800
Kappa	0.467	0.462

ตารางที่ 5 ประสิทธิภาพการทำงานของเทคนิคเพื่อนบ้านใกล้ที่สุดโดยเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ผ่านการสกัดจากวิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม

ตัวชี้วัดประสิทธิภาพ	ชุดข้อมูลต้นฉบับ	วิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม
Accuracy	73.05%	73.59%
Precision	70.84%	71.54%
Specificity	67.75%	68.83%
Sensitivity	78.35%	78.35%
F1 Score	74.40%	74.79%
ROC AUC Score	0.804	0.810
Kappa	0.461	0.472

ตารางที่ 6 ประสิทธิภาพการทำงานของเกรเดียนท์บูตทรีโดยเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ผ่านการสกัดจากวิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่ม

ตัวชี้วัดประสิทธิภาพ	ชุดข้อมูลต้นฉบับ	วิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม
Accuracy	72.77%	73.35%
Precision	72.34%	68.33%
Specificity	71.81%	59.68%
Sensitivity	73.73%	87.02%
F1 Score	73.03%	76.55%
ROC AUC Score	0.801	0.815
Kappa	0.455	0.467

ตารางที่ 7 ประสิทธิภาพการทำงานของป่าสุ่มโดยเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ผ่านการสกัดจากวิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่ม

ตัวชี้วัดประสิทธิภาพ	ชุดข้อมูลต้นฉบับ	วิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม
Accuracy	74.07%	73.68%
Precision	71.76%	71.38%
Specificity	68.76%	68.31%
Sensitivity	79.37%	79.05%
F1 Score	75.37%	75.02%
ROC AUC Score	0.816	0.815
Kappa	0.481	0.474

5 สรุปผลการศึกษา

จากผลการศึกษาข้างต้นจะเห็นได้ว่าประสิทธิภาพการทำงานของแต่ละแบบจำลอง โดยเปรียบเทียบระหว่างชุดข้อมูลต้นฉบับกับชุดข้อมูลที่ผ่านการสกัดจากวิศวกรรมคุณลักษณะร่วมกับเทคนิคป่าสุ่ม มีค่าลดลงแต่ยังคงมีค่าใกล้เคียงกันมากคือป่าสุ่ม และต้นไม้ตัดสินใจ หรือ ในบางแบบจำลองก็มีประสิทธิภาพที่มากกว่าอย่างชัดเจนคือ, เกรเดียนท์บูตทรี, เทคนิคเพื่อนบ้านใกล้ที่สุด และซัพพอร์ตเวกเตอร์แมชชีนโดยจะเห็นว่าคุณลักษณะที่ได้สกัดมานั้นสามารถนำมาสร้างแบบจำลอง และให้ประสิทธิภาพที่เทียบเคียงกับการใช้ชุดข้อมูลต้นฉบับ โดยคุณลักษณะก่อนที่จะทำการสกัดมีทั้งหมด 21 คุณลักษณะ หลังจากใช้การสกัดจากวิศวกรรมคุณลักษณะร่วมกับวิธีป่าสุ่ม ได้ 7 คุณลักษณะดังนี้ 1) ความดันโลหิตสูง 2) สุขภาพโดยทั่วไป 3) ค่าดัชนีมวลกาย 4) อายุ 5) เงินเดือน 6) คลอเรสเตอรอลสูง และ 7) เพศ ดังนั้นการนำวิศวกรรมคุณลักษณะเข้ามาทำงานร่วมกับวิธีป่าสุ่ม ทำให้สามารถคัดกรองคุณลักษณะที่สำคัญที่ส่งผลกระทบต่อการเกิดโรคเบาหวานได้ โดยมีความซับซ้องของการสร้างแบบจำลองลดลงเพราะคุณลักษณะลดลงจาก 21 เหลือเพียง 7 คุณลักษณะ โดยผลลัพธ์ที่ได้สามารถนำไปประยุกต์ใช้สำหรับการประเมินผู้ที่มีแนวโน้มเป็นโรคเบาหวาน รวมไปถึงช่วยแพทย์ในการการตัดสินใจสำหรับการเลือกวิธีการรักษาให้เหมาะสมกับลักษณะของผู้ป่วยได้อย่างรวดเร็วขึ้น นอกจากนี้ผู้ป่วยก็จะไม่เสียโอกาสได้เข้ารับการรักษาตั้งแต่ช่วงระยะแรกที่เริ่มมีแนวโน้มเสี่ยงต่อการเกิดโรคด้วย

กิตติกรรมประกาศ ผู้แต่งขอขอบคุณสาขาวิชาคณิตศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี หลักสูตรนวัตกรรมวิศวชีวการแพทย์ ศูนย์ความเป็นเลิศด้านชีวกลศาสตร์ทางการแพทย์ (Center of Excellence in Biomechanics Medicine) และการได้รับการสนับสนุนจากมหาวิทยาลัยเทคโนโลยีสุรนารีในด้านทุนศักยภาพบัณฑิตในการศึกษาในระดับบัณฑิตศึกษา

เอกสารอ้างอิง

[1] กรมควบคุมโรค. (2564). กรมควบคุมโรค รณรงค์วันเบาหวานโลก ปี 2564 ตระหนักถึงการดูแลรักษาโรคเบาหวาน ให้ได้รับการรักษาอย่างทั่วถึง. สืบค้นจาก https://ddc.moph.go.th/brc/news.php?news=21692&deptcode=brc

[2] MedThai. (2017). *โรคเบาหวาน (Diabetes) อาการ, สาเหตุ, การรักษา, วิธีป้องกัน ฯลฯ.* สืบค้นจาก https://medthai.com/diabetes

[3] เพ็ญพันธุ์ ภูริปัญโญ. (2564). *โรคเบาหวาน (Diabetes mellitus).* สืบค้นจาก https://www.thonburihospital.com/DM.html

[4] โรงพยาบาลพญาไท. (2563). *โรคเบาหวาน คืออะไร*. สืบค้นจาก https://www.phyathai.com/article_detail/2705/th/โรคเบาหวาน_คืออะไร

[5] กอบเกียรติ สระอุบล. (2563). *เรียนรู้ Data Science และ AI : Machine Learning ด้วย Python.* กรุงเทพฯ: หสม มีเดีย เนทเวิร์ค.

[6] PradySin. (2019). *Support Vector Machines (SVM)*. Retrieved from https://medium.com/@pradyasin/support-vector-machines-svm-943f9a732a69

[7] Huawei, (2021). *Machine Learning Algorithms: Decision Trees.* Retrieved from https://forum.huawei.com/enterprise/en/machine-learning-algorithms-decision-trees/thread/710283-895

[8] javatpoint. (2021). *K-Nearest Neighbor (KNN) Algorithm for Machine Learning*. Retrieved from https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

[9] thedatascientist. (2021). *Gradient Boosted Trees* . Retrieved from https://thedatascientist.com/gradient-boosted-trees-python/

[10] Watchapong Daroontham. (2018). *เจาะลึก Random Forest !!!— Part 2 of "รู้จัก Decision Tree, Random Forest, และ XGBoost!!!"*. สืบค้นจาก https://medium.com/@witchapongdaroontham/เจาะลึก-random-forest-part-2-of-รู้จัก-decision-tree-random-forest-และ-xgboost-79b9f41a1c1c

[11] เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิง เบื้องต้น*. กรุงเทพฯ : บริษัท เอเชีย ดิจิตอลการพิมพ์ จำกัด

[12] javaTpoint. (2021). *Feature Engineering for Machine Learning*. Retrieved from https://www.javatpoint.com/feature-engineering-for-machine-learning