การเปรียบเทียบประสิทธิภาพการพยากรณ์ชุดข้อมูลการใช้จักรยานสาธารณะด้วยเทคนิคเหมืองข้อมูล

ธาดา ลิมกุลาคมน์[1] ศุภชัย จตุราวิยสัจ[1] วิกร มาณะวะกริช[1] และไก้รุ่ง เฮงพระพรหม[1*]

[1]สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม

kairung2011.heng@gmail.com[*]

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบเทคนิคการพยากรณ์กลุ่มข้อมูลที่เหมาะสมกับการพยากรณ์ข้อมูลการใช่จักรยานสาธารณะในการวิจัยครังนี้ได้ทำการศึกษาคัดเลือกเทคนิคสำหรับใช่ในการวิจัยนี้ทั้งหมด 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ เพื่อหาประสิทภาพของการพยากรณ์ข้อมูลที่ดีที่สุดในกลุ่มข้อมูล การใช่จักรยานสาธารณะ ผลการวิจัยพบว่า เทคนิคที่ให้ผลลัพธ์ที่ดีสุดคือ การถดถอยเชิงเส้น โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 0.005 +/- 0.009 รองลงมาคือ โครงข่ายประสาทเทียม โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 2.710 +/-0.000 และสุดท้ายคือ ต้นไม้ตัดสินใจ โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 5.768 +/- 0.000

คำสำคัญ : การพยากรณ์ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม ต้นไม้ตัดสินใจ

A Comparison of the Efficiency of Regression Bike Sharing Dataset by Data Mining Technique

Thada Limkulakhom [1] Supachai Jaturawiyasat [1] Wikorn Manawakrit [1] and Kairung Hengpraphrom[1]*

[1] *Data Science Faculty of Science and Technology, Nakhon Pathom Rajabhat University*

**kairung2011.heng@gmail.com*

Abstract

The objective of this research is to study and compare forecasting techniques suitable for data forecasting. Bike Sharing in this research has selected 3 techniques for using in this research, Linear Regression. neural network and decision tree In order to find the best forecasting performance in the Bike of Sharing data group. Techniques that give the best results are Linear regression Where the square root of the mean squared error is 0.005 +/- 0.009, followed by the artificial neural network Where the square root of the mean square mean is 2.710 +/- 0.000 and the final is Decision tree By giving the square root of the mean squared mean value of 5.768 +/- 0.000

Keywords: regression, Linear Regression, neural network, decision tree

1.บทนำ

Bike Sharing คือบริการรถจักรยานสาธารณะที่กำลังได้รับความนิยมในช่วงหลายปีที่ผ่านมานี้ ตามเมืองใหญ่ๆ ของโลกที่ประสบปัญหาด้านการจราจร

คอนเซปง่ายๆ คือ มีจักรยานจอดอยู่ตามท้องถนน อาจจะเป็น Docking Station ที่มีที่ล็อคเป็นระเบียบ แล้วผู้ที่ใช้บริการที่ได้สมัครสมาชิกผ่านระบบและเติมเงิน ปลดล็อคเอาจักรยานไปใช้จาก Station หนึ่ง ขี่ไปยังอีก Station หนึ่ง และทางผู้ให้บริการก็หักเงินตามระยะเวลาที่ใช้ไป

การเดินทางจากจุดหนึ่งไปยังอีกจุดหนึ่ง (Door-to-Door) ด้วยระบบขนส่งสาธารณะ มักประกอบด้วยการเดินทางหลายรูปแบบ (Multimodal Transport) เช่น การนั่งรถมอเตอร์ไซค์รับจ้างจากบ้านไปปากซอย แล้วขึ้นรถไฟฟ้าไป 6 ป้าย จากนั้นจึงเดินเท้าต่อไปอีก 400 เมตร เพื่อไปที่ทำงาน เป็นต้น ในหลายๆ เมืองที่การเดินทางต้องประกอบด้วยหลายรูปแบบ โดยที่บางรูปแบบไม่มีระบบเชื่อมต่อที่มีคุณภาพและประสิทธิภาพ ผู้คนจึงหันมาใช้รถส่วนตัวกันมาก ส่งผลให้เกิดปัญหาด้านการจราจรได้

การใช้จักรยาน Bike Sharing จึงเกิดขึ้นมาเพื่อเป็นระบบเชื่อมต่อกับระบบขนส่งหลัก เช่น ใช้สำหรับเดินทางไปขึ้นรถเมล์ รถไฟ เป็นต้น ส่วนใหญ่การใช้ Bike Sharing จึงเป็นการใช้เพื่อเดินทางระยะสั้นๆ ที่เป็นส่วนหนึ่งของการเดินทาง โดยเรียกว่าเป็น "Last Mile Solution" (การเดินทางที่เชื่อมต่อช่วงสุดท้าย) เช่น การใช้จักรยานขี่จากบ้านไปยังป้ายสถานีรถไฟในตอนเช้า และขี่จากสถานีรถไฟกลับบ้านในตอนเย็น เป็นต้น

แต่ความท้าทายจริงๆ ของระบบนี้อยู่ที่การบริหารจัดการ ไม่ว่าจะเป็นระบบการใช้งาน การซ่อมบำรุงจักรยานให้อยู่ในสภาพที่ใช้งานได้สะดวกและปลอดภัย หรือการจัดการให้มีการกระจายของจักรยานไปตามจุดที่มีคนใช้งานอย่างเพียงพอและเหมาะสม

2.วัตถุประสงค์

เพื่อศึกษาและเปรียบเทียบเทคนิคการพยากรณ์ข้อมูลการใช่จักรยานสาธารณะ ทั้งหมด 3 เทคนิค เทคนิค ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

3.งานวิจัยที่เกี่ยวข้อง

3.1 ทฤษฎีที่เกี่ยวข้อง

ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Decision Tree) หมายถึง ต้นไม้ที่ใช้ในการสนับสนุนการตัดสินใจ ซึ่งมีลักษณะเป็นโครงสร้างต้นไม้หัวกลับที่มีรากอยู่ด้านบนและใบอยู่ด้านล่างสุด โดยที่ภายในต้นไม้จะ ประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดนั้น จะแสดงถึงการตัดสินใจบนข้อมูลของคุณสมบัติต่างๆ กิ่งของต้นไม้แสดงถึงค่าหรือผลลัพธ์ที่ได้จากการทดสอบ และใบซึ่งเป็นสิ่งที่

อยู่ล่างสุดของต้นไม้ตัดสินใจจะแสดงถึงกลุ่มของข้อมูล(Class) หรือผลลัพธ์โหนดที่อยู่บนสุดจะเรียกว่า โหนดราก (Root Node) (นฤพนต์ ว่องประชานุกลู , 2548)

โครงข่ายประสาทเทียม (Artificial Neural Networks - ANN)

โครงข่ายประสาทเทียม (Artificial Neuron Network : ANN) เป็นศาสตร์แขนงหนึ่งของทางด้านปัญญาประดิษฐ์ (Artifitial Intelligence : AI) มีรูปแบบโครงสร้างและการทำงานของการประมวลผลเหมือนกับสมองของสิ่งมีชีวิตซึ่งมีปรับเปลี่ยนตัวเองต่อการตอบสนองของอินพุตตามกฎของการเรียนรู้ (learning rule) หลังจากที่โครงข่ายได้เรียนรู้สิ่งที่ต้องการแล้ว โครงข่ายนั้นจะสามารถทำงานที่กำหนดไว้ได้โครงข่ายประสาทเทียมได้ถูกพัฒนาคิดค้นจากการทำงานของสมองมนุษย์โดยสมองมนุษย์ประกอบไปด้วยหน่วยประมวลผลเรียกว่า นิวรอน (เซลล์ประสาท หรือ neuron) จำนวนนิวรอลในสมองมนุษย์มีอยู่ประมาณและมีการเชื่อมต่อกันอย่างมากมาย สมองมนุษย์จึงสามารถกล่าวได้ว่าเป็นคอมพิวเตอร์ที่มีการปรับตัวเอง (adaptive) ไม่เป็นเชิงเส้น (nonlinear) และทำงานแบบขนาน (parallel) ในการดูแลจัดการการทำงานร่วมกันของนิวรอนในสมอง การคำนวณเชิงนิวรอลเป็นการคำนวณที่เลียนแบบมาจากการทำงานของสมองมนุษย์นั่นเอง

การถดถอยเชิงเส้น (linear regression)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ซึ่งได้แก่ตัว ประมาณการ (Predictor, X) และตัวตอบสนอง (Response, y) โดยเป็นความสัมพันธ์แบบเชิงเส้น (Linear) ทั้งนี้ในขั้นตอนการทำ Regression ต้องมีการเก็บจำนวน Sample space จำนวนมากพอ นั้นคือ มี x และ y ที่มีความสัมพันธ์กันหลายๆ ครั้ง เพื่อนำมาหาสมการความสัมพันธ์

3.2 งานวิจัยที่เกี่ยวข้อง

การพยากรณ์ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตโดยใช้วิธีวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์ปริมาณการชำระเงินผ่านบริการ ธนาคารบนอินเทอร์เน็ตของธนาคารพาณิชย์ในประเทศไทย โดยใช้วิธีการวิเคราะห์อนุกรมเวลาด้วย

เทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียมแบบ เปอร์เซ็ปตรอนหลายชั้น

และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย

ข้อมูลอนุกรมเวลาที่นำมาใช้ในการศึกษาเป็นข้อมูลปริมาณการชำระเงินผ่านบริการธนาคารบนอินเทอร์เน็ต ตั้งแต่ปี 2553 ถึง 2558

จากผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณการชำระเงินผ่านบริการธนาคารบนอินเทอร์เน็ตกับชุดข้อมูลในปี 2558 พบว่าแบบจำลองการพยากรณ์ด้วยเทคนิควิธี โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีประสิทธิภาพโดยรวมสูงที่สุด โดยมีค่าเฉลี่ยความ คลาดเคลื่อนสัมพัทธ์เท่ากบั 8.70% ลำดับถัดมาได้แก่ แบบจำลองการพยากรณ์ด้วยเทคนิควิธีการถดถอย เชิงเส้น โดยมีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 16.75%

การพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซลโดยเทคนิควิธีเหมืองข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนา และวัดประสิทธิภาพของแบบจำลองการพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซล โดยเทคนิควิธีเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย

ข้อมูลที่นำมาใช้ศึกษาเป็นข้อมูลอนุกรมเวลารายเดือนปริมาณ การใช้น้ำมันดีเซล เผยแพร่โดยกระทรวงพลังงาน โดยใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม 2549 ถึงวันที่ 31 ธันวาคม 2561 ซึ่งผู้วิจัยได้ทำการแบ่งชุดข้อมูล 2 ชุดสำหรับสร้างแบบจำลองจำนวน 144 รายการ และสำหรับทดสอบแบบจำลอง 12 รายการจากการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซล ที่มี ความเหมาะสมกับชุดข้อมูลในปี 2561

พบว่า แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีประสิทธิภาพ โดยรวมสูงที่สุด มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 2.38% รองลงมาคือ แบบจำลองการพยากรณ์ด้วยเทคนิคการ ถดถอยเชิงเส้น มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 2.66%

การพยากรณ์จำนวนผู้ป่วยในโรคหลอดเลือดหัวใจที่จำหน่ายออกจาก โรงพยาบาล ในคณะแพทยศาสตร์วชิรพยาบาล ด้วยวิธีการวิเคราะห์ อนุกรมเวลาโดยใช้เทคนิคปรับให้เรียบเอกซ์โพเนนเชียลแบบง่าย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์จำนวนผู้ป่วยในโรคหลอดเลือดหัวใจที่จำหน่ายออกจากโรงพยาบาลรายปีงบประมาณ 2559 -2563 ในคณะแพทยศาสตร์วชิรพยาบาลโดยใช้เทคนิคปรับให้เรียบเอกซ์โพเนนเชียลแบบง่าย

วิธีการดำเนินงานวิจัย การศึกษานี้เป็นรูปแบบติดตามไปข้างหน้าโดยใช้ข้อมูลในอดีต เพื่อพยากรณ์จำนวนผู้ป่วย ในโรคหลอดเลือดหัวใจที่คาดว่าจะจำหน่ายรายปีงบประมาณ 2559 - 2563 กลุ่มตัวอย่างคือ จำนวนผู้ป่วย ในโรคหลอดเลือดหัวใจ ICD-10 รหัส I25 ที่จำหน่ายรายปีของปีงบประมาณ 2549 - 2558 เก็บรวบรวมข้อมูล จากสถิติผลการปฏิบัติงานปีงบประมาณ 2549-2558 ของคณะแพทยศาสตร์วชิรพยาบาล มหาวิทยาลัย นวมินทราธิราช

สำหรับดัชนีที่ใช้ประเมินความแม่นยำของการพยากรณ์ในการศึกษานี้ ได้แก่ ค่าเบี่ยงเบน สัมบูรณ์เฉลี่ย (Mean Absolute Deviation, MAD), ค่าเปอร์เซ็นต์ผิดพลาดเฉลี่ยสัมบูรณ์ (Mean Absolute Percent Error, MAPE) และะค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error; MSE) ที่มีค่าต่ำที่สุด

แนวโนม้จำนวนผู้ป่วยยในโรคหลอดเลือดหัวใจที่คาดว่าจะจำหน่ายรายปีงบประมาณ 2559 ประมาณอย่างน้อยเท่ากับ 608 รายต่อปีงบประมาณ โดยกำหนดค่าคงที่ปรับเรียบ หรือค่า α เท่ากับ 1 จะได้ค่าความคลาดเคลื่อน ของตัวแบบพยากรณ์ต่ำที่สุด ได้แก่ ค่า MAD เท่ากับ 162.78 ค่า MAPE เท่ากับ 19.01% และค่า MSE เท่ากับ 36749.67 ซึ่งถือว่าความแม่นยำของการพยากรณ์อยู่ในระดับดี และในปีงบประมาณ 2560-2563 อย่างน้อย 608 รายต่อปีงบประมาณ มีค่าความคลาดเคลื่อนที่ต่ำสุดคือ ค่า MAD เท่ากับ 104.64, ค่า MAPE เท่ากับ 12.22% และค่า MSE เท่ากับ 23624.79 ในปีงบประมาณ 2563

การเปรียบเทียบวิธีการพยากรณ์สำหรับความต้องการใช้พลังงานไฟฟ้าในจังหวัดนครพนม

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการพยากรณ์ สำหรับการพยากรณ์ความต้องการใช้พลังงานไฟฟ้าในจังหวัด นครพนม โดยใช้ข้อมูลจากการไฟฟ้าฝ่ายผลิตจังหวัดนครพนม ตั้งแต่เดือนมกราคม 2554 ถึงเดือนกันยายน 2559 จำ นวน 69 ค่า โดยแบ่ง ข้อมูลออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่เดือนมกราคม 2554 ถึงเดือนธันวาคม 2558 จา นวน 60 ค่า สา หรับการเปรียบเทียบหาวิธีการ พยากรณ์ที่เหมาะสมที่สดุ โดยใช้เกณฑ์พิจารณาค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (MAD) และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ที่ต่ำที่สดุ

งานวิจัยนี้ใช้วิธีพยากรณ์ 6 วิธีคือ วิธีค่าเฉลี่ยพื้นที่ วิธีแนวโน้มเชิงเส้น วิธปีรับเรียบเอก็ซ์โพเนนเชียลอย่างง่าย วิธีปรับเรียบเอก็ซ์โพเนนเชียลแบบโฮลท์ วิธีปรับเรียบเอก็ซ์โพเนนเชียลแบบวินเทอร์ และวิธีแยกส่วนประกอบ

จากนั้นจึง เลือกวิธีการที่พยากรณ์ที่เหมาะสมที่สุดมาคำนวณหาช่วงการพยากรณ์ล่วงหน้ากับข้อมูลชุดที่ 2 คือข้อมูลตั้งแต่เดือนมกราคม 2559 ถึงเดือน กันยายน 2559 จา นวน 9 ค่าโดยใช้เกณฑ์พิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ที่ต่ำ ที่สุด

ผลการศึกษาพบว่า วิธีการพยากรณ์ที่มีความเหมาะสมที่สดุคือการพยากรณ์โดยวิธีแยกส่วนประกอบ จากรปูแบบดังกล่าวนำมาคำนวณหาช่วงพยากรณ์ล่วงหน้า 3 เดือน 6 เดือน และ 9 เดือน พบว่าวิธีนี้เหมาะสำหรับการพยากรณ์ล่วงหน้า 9 เดือน

ตัวแบบพยากรณ์ราคามังคุดคละ

วัตถุประสงค์ของการวิจัยครั้งนี้ คือ การสร้างตัวแบบพยากรณ์ที่เหมาหรับพยากรณ์ราคาะสม มังคุดคละที่เกษตรกรขายได้ ณ ไร่นาโดยใช้ข้อมูลจากเว็บไซต์ของสำนักงานเศรษฐกิจการเกษตั้งแต่เดือนร มกราคม 2548 ถึงเดือนสิงหาคม 2558 จำนวน 128 ค่าผู้วิจัยได้แบ่งข้อมูลออกเป็น2ชุดชุดที่1คือ ข้อมูลตั้งแต่ เดือนมกราคม 2548 ถึงเดือนธันวาคม 2557จำนวน 120ค่าสำหรับการสร้างตัวแบบพยากรณ์

วิธีการพยากรณ์ที่ใช้ในงานวิจัยมี 4 วิธี คือ วิธีการทาง สถิติ3 วิธี ได้แก่ วิธีบอกซ์-เจนกินส์ วิธีการแยกส่วนประกอบอนุกรมเวลา และวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้

กำลังของวินเทอร์แบบคูณชุดที่2คือ ข้อมูลตั้งแต่เดือนมกราคมถึงเดือนสิงหาคมจำนวน2558 8 ค่าสำหรับการตรวจสอบความแม่นของตัวแบบพยากรณ์ด้วยเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด ผลการวิจัยพบว่า จากวิธีารพยากรณ์ทั้งหมดที่ได้ศึกษาวิธีการแยกส่วนประกอบอนุกรมเวลาเป็นวิธีที่มีความ เหมาะสมกับอนุกรมเวลาชุดนี้มากที่สุดตรวจสอบความแม่นของตัวแบบพยากรณ์ด้วยเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด

ผลการวิจัยพบว่า จากวิธีารพยากรณ์ทั้งหมดที่ได้ศึกษาวิธีการแยกส่วนประกอบอนุกรมเวลาเป็นวิธีที่มีความเหมาะสมกับอนุกรมเวลาชุดนี้มากที่สุด

การศึกษาและเปรียบเทียบวิธีการพยากรณ์ สำหรับการพยากรณ์ความต้องการพลังไฟฟ้าสูงสุดในเขตภาคกลาง ของประเทศไทย

การวิจัยนี้เป็นการศึกษาและเปรียบเทียบวิธีการพยากรณ์ สำหรับการพยากรณ์ความต้องการพลังไฟฟ้าสูงสุดในเขตภาคกลาง ของประเทศไทย โดยแบ่งข้อมูลออกเป็น 2ส่วนที่ข้อมูล1ตังแต่เดือนมกราคม 2545 ถึง เดือน ธันวาคม 2550 สำหรับกำหนดรูปแบบและสวนที่ข้อมูลตั้งแต่เดือนมกราคม 2551 ถึง เดือนธันวาคม

วิธีการพยากรณ์ที่ใช้ในงานวิจัยมี 4 วิธี คือ วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียล แบบ Holt-Winter การวิเคราะห์การถดถอยที่ใช้ตัวแปรดัมมี่ วิธีของบอกซ์-เจนกินส์ และการวิเคราะห์การถดถอยแบบ ฟัซซีที่ใช้ตัวแปรดัมมี่

การเลือกวิธีการพยากรณ์ที่เหมาะสมพิจารณาจากคMAPEที่ต่ำที่สุดผลการศึกษา พบว่าวิธีการ พยากรณ์ที่ให้รูปแบบที่เหมาะสมที่สุด คือ วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลแบบHolt-Winterจากรูปแบบดังกล่าว นำมาคำนวณช่วงการพยากรณ์ล่วงหน้า2 6 และ12 เดือน พบว่าวิธีนี้เหมาะสำหรับการพยากรณ์ระยะสั้นล่วงหน้า 2เดือน

4.ขั้นตอนการดำเนินงาน

4.1 ข้อมูลสำหรับใช้ในการทดลอง

ผู้วิจัยได้ทำการรวบรวมข้อมูลการใช่จักรยานสาธารณะ โดยนำข้อมูลดังกล่าวมาจากเว็บไซต์ MLData จากนั้นนำข้อมูลมาทำความสะอาดโดยตัดข้อมูลที่มีค่าขาดหาย (Missing Values) ออก

4.2 การดำเนินการทดลอง

1) นำข้อมูลที่ได้จัดเตรียมไว้ในขั้นตอนที่ 4.1 มาทำการเปรียบเทียบประสิทธิภาพการพยากรณ์การใช่จักรยานสาธารณะ ด้วยเทคนิคการพยากรณ์ จำนวน 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

2) นำผลลัพธ์จากการจำแนกกลุ่มที่ได้จากการทดลองมาหาประสิทธิภาพการพยากรณ์การใช่จักรยานสาธารณะ

3) นำเทคนิคการพยากรณ์กลุ่มข้อมูลการใช่จักรยานสาธารณะที่มีประสิทธิภาพมากที่สุดไปพัฒนาระบบการพยากรณ์การใช่จักรยานสาธารณะต่อไป

5.ผลการดำเนินงาน

ผลการทดลองเปรียบเทียบประสิทธิภาพในการการพยากรณ์การใช่จักรยานสาธารณะ จำนวน 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ

ตารางที่ 1 ผลการทดลองเมื่อเปรียบเทียบทั้ง 3 เทคนิค

Dataset	
Method	Root mean square error
Linear Regression	0.005 +/- 0.009
neural network	2.710 +/- 0.000
decision tree	5.768 +/- 0.000

ภาพที่ 1 ผลการทดลองเมื่อเปรียบเทียบทั้ง 3 เทคนิค

ผลการทดลองการเปรียบเทียบประสิทธิภาพการพยากรณ์ข้อมูลการใช่จักรยานสาธารณะ 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ ผลการวิจัยพบว่า เทคนิคที่ให้ผลลัพธ์ที่ดีสุดคือ การถดถอยเชิงเส้น โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 0.005 +/- 0.009 รองลงมาคือ โครงข่ายประสาทเทียม โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 2.710 +/-0.000 และสุดท้ายคือ ต้นไม้ตัดสินใจ โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 5.768 +/- 0.000

6.สรุปผลการดำเนินงานและข้อเสนอแนะ

6.1 สรุปผลการดำเนินงาน

งานวิจัยนี้ได้ทำการศึกษาถึงประสิทธิภาพของวิธีการพยากรณ์ข้อมูลการใช่จักรยานสาธารณะด้วยเทคนิคการพยากรณ์ข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และต้นไม้ตัดสินใจ ผลการวิจัยพบว่า เทคนิคทั้ง 3 เทคนิค เทคนิคที่ให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสองน้อยที่สุด จะมีประสิทธิภาพดีที่สุด ดั้งนั้นเทคนิคที่ให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสองน้อยที่สุด คือ การถดถอยเชิงเส้น โดยให้ค่ารากที่สองของค่าเฉลี่ยความผิดผลาดกำลังสอง เท่ากับ 0.005 +/- 0.009

6.2 ข้อเสนอแนะ

เนื่องจากเทคนิคการพยากรณ์มีหลากหลายเทคนิค ซึ่งเทคนิคที่เลือกมาเปรียบเทียบ 3 เทคนิคนั้นอาจจะไม่ใช่วิธีการที่ดีที่สุด ดังนั้น ควรทดลองใช้วิธีการอื่น ๆ ที่ยังไม่ได้ถูกนำมาทำการเปรียบเทียบการพยากรณ์ข้อมูลได้หากนำทุกวิธีการมาทำการเปรียบเทียบกันทั้งหมด อาจจะได้วิธีการพยากรณ์ข้อมูลการใช่จักรยานสาธารณะที่ดีกว่า

7. เอกสารอ้างอิง

ณรงค์ฤทธิ์ สุคนธสิงห์. (2559). **การพยากรณ์ปริมาณการชำระเงินผ่านธนาคารบนอินเทอร์เน็ตโดยใช้วิธีวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล** คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม.

อรรณพ กางกั้น. (2562). **การพยากรณ์อนุกรมเวลาปริมาณการใช้น้ำมันดีเซลโดยเทคนิควิธีเหมืองข้อมูล** การประชุมวิชาการเสนอผลงานวิจัยระดับชาติ ด้านวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏจันทรเกษม ครั้งที่ 3. กรุงเทพมหานคร.

สุมิตรา เมืองขวา. (2559). การพยากรณ์จำนวนผู้ป่วยในโรคหลอดเลือดหัวใจที่จำหน่ายออกจากโรงพยาบาล. **วชิรเวชสารและวารสารเวชศาสตร์เขตเมือง มหาวิทยาลัยนวมินทราธิราช.** 60(4): 277-286.

เฉลิมชาติ ธีระวิริยะ. (2560). การเปรียบเทียบวิธีการพยากรณ์สำหรับความต้องการใช้พลังงานไฟฟ้าในจังหวัดนครพนม. **วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยณเรศวร.** 25(4): 124-137.

วรางคณา เรียนสุทธิ์ (2560) ตัวแบบพยากรณ์ที่เหมาะหรับพยากรณ์ราคาะมังคุดคละ. **วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี.** 19(2): 31-42.

จินตพร หนิ้วอินปั๋น บุญอ้อม โฉมที และประสิทธิ์ พยัคฆพงษ์. (2555). **การเปรียบเทียบวิธีการพยากรณ์ 4 วิธีสำหรับความต้องการพลังไฟฟ้าสูงสุด ในภาคกลางของประเทศไทย** ประชุมวิชาการเสนอผลงานวิจัย ระดับบัณฑิตศึกษา ครั้งที่ 13 มหาวิทยาลัยขอนแก่น.