

การทำนายจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทย
โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล
**Forecasting International Tourist Arrival to Thailand by
Nationality Using Time Series Data Mining Techniques**

พิชญากร เลค^{1*}
Pitchayakorn Lake^{1*}

ABSTRACT

This research aim to develop the model of forecasting the number of International Tourists arrival to Thailand. They can expand economic of Thailand. The research provide 3 techniques, which techniques are suited to the forecasting model i.e. Linear Regression, Multi-Layer Perceptron and Support Vector Machine for Regression. The study data was the number of the International Tourists arrival from 2007 to 2015, which they were on the amount of 108 months.

The experimental results show that the forecasting trends the number of ASEAN travelers to visit Thailand i.e. China, Malaysia, Japan and Korea. The forecasting model achieves the highest accuracy rate as follows, 1) the support vector machine for regression forecasting was the most achievable for China, Malaysia and Korea, which the highest accuracy rate were 12.07%, 8.41% and 14.29% of MMRE, respectively. 2) the linear regression forecasting was the most achievable for Japan, which the highest accuracy rate was 8.17% of MMRE.

Keywords: Forecasting, Number of International Tourist Arrival to Thailand, Time Series Analysis, Data Mining Techniques

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์ปริมาณนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย เพื่อพัฒนาเศรษฐกิจด้านการท่องเที่ยวให้เติบโตขึ้น โดยใช้วิธีการวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิคได้แก่ 1) การถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยข้อมูลที่น่ามาใช้ศึกษาเป็นข้อมูลนักท่องเที่ยวชาวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2550 – 2558 จำนวน 108 เดือน

จากผลการวิจัยพบว่าแบบจำลองที่มีความเหมาะสมในการทำนายแนวโน้มจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทย ในกลุ่มทวีปเอเชีย ได้แก่ ประเทศจีน ประเทศมาเลเซีย ประเทศญี่ปุ่น และประเทศเกาหลี คือแบบจำลองการพยากรณ์ด้วยซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีความเหมาะสมกับชุดข้อมูลประเทศจีน มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 12.07% แบบจำลองการพยากรณ์ด้วยซัพพอร์ตเวกเตอร์

¹ ภาควิชาคอมพิวเตอร์ธุรกิจ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม กรุงเทพฯ 10160

Department of Business Computer, Faculty of Information Technology, Siam University, Bangkok 10160, Thailand.

*Corresponding author: E-mail pitchayakorn@siam.edu

แมชชีนสำหรับการถดถอยมีความเหมาะสมกับชุดข้อมูลประเทศมาเลเซีย มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 8.41% แบบจำลองการพยากรณ์ด้วยวิธีการถดถอยแบบเชิงเส้น มีความเหมาะสมกับชุดข้อมูลประเทศญี่ปุ่น มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 8.17% และแบบจำลองการพยากรณ์ด้วยซัพพอร์ตเวกเตอร์แมชชีน สำหรับการถดถอย มีความเหมาะสมกับชุดข้อมูลประเทศเกาหลี มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 14.29%

คำหลัก: ปริมาณนักท่องเที่ยว, การวิเคราะห์อนุกรมเวลา, เทคนิคเหมืองข้อมูล

คำนำ

การท่องเที่ยวเป็นรายได้สำคัญที่ทำให้เศรษฐกิจของประเทศเจริญเติบโต ในปีที่ผ่านมา ปริมาณนักท่องเที่ยวชาวต่างชาติลดลงอย่างต่อเนื่อง บางครั้งสถานการณ์ภายในประเทศมีความสงบ จนได้รับความเชื่อมั่นจากชาวต่างชาติเพิ่มขึ้นอย่างต่อเนื่อง เนื่องจากปัญหาทางด้านเศรษฐกิจ ปัญหาการเมืองและปัญหาการก่อการร้ายภายในประเทศของนักท่องเที่ยวต่างประเทศเอง ทำให้ปริมาณนักท่องเที่ยวมีแนวโน้มลดลง

กระทรวงการท่องเที่ยวและกีฬา ร่วมกับหน่วยงานและภาคต่าง ๆ ได้มีการรณรงค์จัดกิจกรรมต่าง ๆ เพื่อส่งเสริมการท่องเที่ยว ซึ่งสามารถกระตุ้นการขยายตัวของนักท่องเที่ยว นอกจากนี้ศูนย์วิจัยกสิกรพบว่า การที่ภาครัฐบาลมีนโยบายที่เน้นนักท่องเที่ยวต่างชาติกลุ่มคุณภาพ และสร้างรายได้จากการท่องเที่ยวให้กระจายไปสู่ชุมชนในท้องถิ่นเป็นนโยบายที่เพิ่มรายได้ให้แก่การท่องเที่ยวในประเทศไทย การใช้นโยบายนี้ควรมีการพัฒนาผลิตภัณฑ์ในการท่องเที่ยวให้มีมาตรฐาน เพื่อกระจายรายได้จากนักท่องเที่ยวไปสู่ชุมชนท้องถิ่น เช่น การจัดตั้งศูนย์จำหน่ายสินค้าโอท็อป (OTOP) ท้องถิ่นที่ได้มาตรฐานและปลอดภาษี (Duty Free OTOP) หรือการขอคืนภาษี (VAT Refund) ในกรณีซื้อสินค้ากับร้านค้าที่จดทะเบียนอย่างถูกต้อง เป็นต้น

งานวิจัยฉบับนี้ ผู้วิจัยเลือกชุดข้อมูลนักท่องเที่ยวชาวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2550 – 2558 ซึ่งเป็นชุดข้อมูลที่รัฐบาลได้นำมาเผยแพร่ ผ่านทางเว็บไซต์ <http://www.nso.go.th> คือ เว็บไซต์สำนักงานเศรษฐกิจ

แห่งชาติ สังกัดสำนักนายกรัฐมนตรี ในหน้าแรก ประกอบด้วย คู่มือการใช้งาน โครงการต่างๆ แบ่งเป็น 3 ด้าน ได้แก่ ด้านประชากรและเคหะ ด้านเศรษฐกิจ ด้านพลังงานและสิ่งแวดล้อม หัวข้อย่อยบริการด้านสถิติแบ่งเป็น สืบค้นตารางสถิติ สถิติอนุกรมเวลา (Time Series) มุมมองเมืองไทยด้วยสถิติ บทความเด่น สถิติ 10 อันดับเด่น รายได้และความยากจน หัวข้อย่อย ความคิดเห็นของประชาชน หัวข้อย่อย สิ่งพิมพ์ ได้แก่ เศรษฐกิจ สังคม ไอซีทีเอกสารรายงานสถิติ e-Book สารสถิติ รายงานประจำปี เป็นต้น เพื่อให้สอดคล้องกับสถานการณ์และมีความชัดเจนในการดำเนินการในฐานะหน่วยงานกลางด้านงานสถิติของรัฐบาล เพื่อความคล่องตัวในการผลิตสถิติให้ทันสถานการณ์ เป็นประโยชน์ต่อการนำไปใช้ในการพัฒนาและแก้ไขปัญหาในการบริหารงานของรัฐบาลมากขึ้น

จึงเห็นได้ว่าแนวทางในการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย มีหลายวิธีที่สามารถวิเคราะห์และประเมินจำนวนนักท่องเที่ยวได้ สำหรับงานวิจัยนี้ผู้วิจัยเลือกใช้วิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น (Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) โดยผลลัพธ์การประมวลผลข้อมูลอนุกรมเวลา อยู่ในรูปของแบบจำลองพยากรณ์ปริมาณนักท่องเที่ยว สำหรับชุดข้อมูลตัวอย่าง สำหรับการทดลองในการวิจัย ผู้วิจัยเลือกประเทศ 4 ประเทศ

ที่มีปริมาณนักท่องเที่ยวต่างประเทศมีแนวโน้มในการมาท่องเที่ยวในประเทศไทยเพิ่มขึ้น เพื่อสนับสนุนให้รัฐบาลมีนโยบายจัดกิจกรรมต่าง ๆ เพื่อสนับสนุนจำนวนนักท่องเที่ยวให้ขยายตัวเพิ่มขึ้น เป็นการกระตุ้นเศรษฐกิจ นำรายได้เข้าประเทศและกระจายรายได้สู่ชุมชนและสถานประกอบการในท้องถิ่นเพิ่มขึ้น

วัตถุประสงค์ในการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพ และเลือกเทคนิคที่เหมาะสมในการวิเคราะห์อนุกรมเวลาของจำนวนนักท่องเที่ยวต่างประเทศ ที่เดินทางเข้ามาในประเทศไทย
2. เพื่อสร้างรูปแบบการพยากรณ์จำนวนนักท่องเที่ยวที่มีแนวโน้มในการเดินทางมาประเทศไทยเพิ่มขึ้น ด้วยวิธีวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล

เอกสารงานวิจัยที่เกี่ยวข้อง

1. การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ซึ่งในปัจจุบัน การทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ด้านวิทยาศาสตร์การแพทย์ ด้าน เศรษฐกิจ และ สังคม (Rutherford, 2011) คอมพิวเตอร์สามารถเรียนรู้และจดจำข้อมูลต่าง ๆ เพื่อใช้ในการค้นหารูปแบบ (Pattern) กฎเกณฑ์ (Rule) ของความสัมพันธ์ (Data Set) เพื่อสกัดข้อมูลซึ่งอยู่ในรูปแบบของสารสนเทศ (Information) เพื่อใช้ในกระบวนการสร้างองค์ความรู้ และค้นหาค้นหาองค์ความรู้ที่มีอยู่ในข้อมูล (Knowledge Discovery in Database: KDD) เหมืองข้อมูลถือว่าเป็นการวิเคราะห์เชิงลึก ผลลัพธ์และกระบวนการขึ้นอยู่กับวัตถุประสงค์ ดังนั้น ผู้ใช้เหมืองข้อมูลต้องมีความรู้หลายด้าน เพื่อหาผลลัพธ์ในการวิเคราะห์ที่ดีที่สุด (Wu, 2004)

2. การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

คือ วิธีการใช้ข้อมูลในอดีตเพื่อพยากรณ์ข้อมูลในอนาคต โดยข้อมูลในอดีตจะเก็บรวบรวมเป็น รายวัน รายสัปดาห์ รายเดือน หรือรายปีอย่างต่อเนื่อง แบ่งออกเป็น 2 กลุ่ม ได้แก่

2.1 กลุ่มค่าเฉลี่ยเคลื่อนที่ (Moving Average) วิธีการนี้เหมาะกับข้อมูลที่มีลักษณะคงที่ไม่เปลี่ยนแปลงมากในแต่ละงวด เช่น ยอดขายของสินค้าหรือบริการที่ไม่มีอิทธิพลของฤดูกาล (Seasonal) มาเกี่ยวข้อง

2.2 กลุ่มเทคนิคปรับเรียบเส้นโค้ง (Smoothing Technique) เป็นวิธีที่เหมาะสมกับข้อมูลที่ไม่ค่อยเปลี่ยนแปลง ใช้หลักการเดียวกับค่าถ่วงเฉลี่ยแบบเคลื่อนที่อย่างง่าย คือใช้ข้อมูลในอดีตมาถ่วงน้ำหนัก แต่น้ำหนักที่ถ่วงข้อมูลกับข้อมูลในอดีตไม่เท่ากัน (อัจฉรา, 2544)

3. การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression)

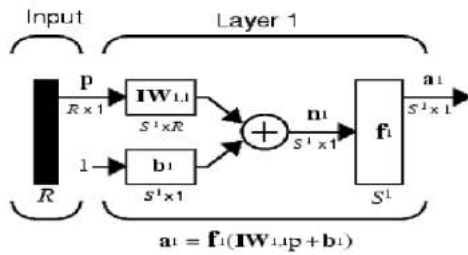
เป็นวิธีการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ ซึ่งเทคนิคนี้จะอาศัยความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรเพื่อใช้ทำนาย Cai *et al.* (2006) โดยสมการ Linear Regression จะแสดงให้เห็นค่าความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระในลักษณะของข้อมูลเชิงปริมาณ โดยสามารถเขียนได้ดังนี้

$$Y = a + bx_1 + cx_2 + \dots \quad (1)$$

4. โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron : MLP)

โครงข่ายประสาทเทียมแบบหลายชั้น (Kubat, 1999) ที่มีลักษณะการเชื่อมต่อแบบโยงไปข้างหน้าแบบทั่วถึง (Fully connected feed-forward nets) สามารถมีจำนวนชั้นตั้งแต่หนึ่งชั้นขึ้นไป Martinez *et al.* (2006) ถูกพัฒนาขึ้นมาเพื่อแก้จุดอ่อนของโครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Perceptron) ให้ความสามารถในการคำนวณที่สูงขึ้น โดยโครงข่ายประสาทเทียมแบบหลายชั้นจะประกอบไปด้วย ชั้นข้อมูลเข้า (Input Layer) ตามจำนวนแอททริบิวต์ของชุดข้อมูล ชั้นซ่อน (Hidden Layer) ความ

เหมาะสมของการกำหนดขึ้นอยู่กับทดสอบประสิทธิภาพ และชั้นผลลัพธ์ (Output Layer)



รูปที่ 1 โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron: MLP)

5. ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression)

ซัพพอร์ตเวกเตอร์แมชชีน หรือ SVM เป็นอัลกอริทึมที่ใช้ในการจำแนกกลุ่มของข้อมูลด้วยวิธีการหาระนาบการตัดสินใจ (Decision Hyper Plane) หรือไฮเปอร์เพลนที่เหมาะสม สำหรับการแบ่งข้อมูล 2 ส่วนจากกัน SVM ยังถูกนำมาประยุกต์ใช้ในการสร้างสมการในการประมาณการหาค่าฟังก์ชันเชิงเส้น $f(x)$ ในการใช้ระนาบตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย หรือ SVR จะเป็นการนำข้อมูลปัจจุบัน และข้อมูลในอดีตจำนวนหนึ่งมาทำการเรียนรู้ (Training) เพื่อให้ทราบรูปแบบสำหรับคาดการณ์ผลซึ่งจะเกิดขึ้นในอนาคต ในงานวิจัยนี้ผู้วิจัยใช้วิธีสร้างสมการค่าฟังก์ชันเชิงเส้นด้วยวิธี Sequential Minimal Optimization for SVM Regression (SMOreg) Shevade *et al.* (2000), (Smola and Schölkopf, 2004)

6. รากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)

เป็นค่าที่ใช้ในการวัดขนาดของความคลาดเคลื่อนของการพยากรณ์ โดยค่าดังกล่าวได้จากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) ซึ่งเป็นการนำผลต่างของค่าจริงและค่าที่ได้จากการพยากรณ์มายกกำลังสอง อย่างไรก็ตามถ้าค่าผลต่างมีค่ามาก จะส่งผลให้ค่าความคลาดเคลื่อนมีค่าที่สูง จึงมีการนำค่าดังกล่าวมาคำนวณด้วยรากที่สอง (Square Root) เพื่อให้ค่าดังกล่าวมีหน่วยวัด

เดียวกับค่าที่ทำกรทดลอง สำหรับค่า RMSE ที่ได้จากการทดลองมีค่าน้อย จะแสดงให้เห็นว่าตัวแบบการพยากรณ์ สามารถทำนายผลลัพธ์ที่มีความคลาดเคลื่อนที่ต่ำ หรือกล่าวได้ว่าแบบจำลองการพยากรณ์มีประสิทธิภาพที่ดี ค่า RMSE สามารถคำนวณได้จากสมการที่ (2)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2)$$

เมื่อ Y_i คือ ค่าประมาณการจากแบบจำลองการพยากรณ์ \hat{Y}_i คือ ค่าจริงที่ใช้ในการสร้างแบบจำลอง n คือ จำนวนข้อมูลในชุดข้อมูล

7. ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE)

เป็นค่าเฉลี่ยของความแตกต่างสัมบูรณ์ระหว่างค่าพยากรณ์และค่าจริง หากค่า MAE มีค่าน้อย แสดงว่าแบบจำลองสามารถประมาณค่าประมาณได้ใกล้เคียงกับค่าจริง ค่า MAE สามารถคำนวณได้จากสมการที่ (3)

$$MAE = \frac{1}{N} \sum_{i=1}^N (T_i - F_i) \quad (3)$$

เมื่อ T_i คือ ค่าจริง F_i คือ ค่าพยากรณ์ N คือ จำนวนข้อมูลในชุดข้อมูล

อุปกรณ์และวิธีการ วิธีดำเนินการวิจัย

1. ศึกษาปัญหาและวิเคราะห์ข้อมูล

รายได้จากการท่องเที่ยวเป็นรายได้จำนวนมากมหาศาล สามารถนำไปพัฒนาประเทศในด้านต่าง ๆ ปีที่ผ่านมานักท่องเที่ยวเดินทางมาในประเทศไทยมีจำนวนมาก แต่บางประเทศนักท่องเที่ยวเดินทางเข้ามาในประเทศไทยมีจำนวนน้อยลง อันเนื่องมาจากสถานการณ์ไม่สงบทางการเมืองภายในประเทศนั้น ดังนั้นรัฐบาลควรมีนโยบายสนับสนุนกิจกรรมต่าง ๆ ด้านการท่องเที่ยวให้ขยายตัวเพิ่มขึ้น การฟื้นตัวของตลาดนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย สะท้อนให้เห็นได้ว่า ประเทศไทยสามารถเป็นจุดหมายปลายทางของนักท่องเที่ยว

หลายประเทศ ซึ่งการดูแลภาคการท่องเที่ยว นั้น เป็นกลไกในการขับเคลื่อนเศรษฐกิจของประเทศไทยในปีต่อไปในอนาคต อย่างไรก็ตามนักท่องเที่ยวบางกลุ่มยังคงลดลงอย่างต่อเนื่อง ผู้วิจัยจึงพัฒนางานวิจัยนี้เพื่อวัตถุประสงค์ในการสร้างแบบจำลองการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย โดยอาศัยเทคนิควิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิควิธีเหมืองข้อมูล เพื่อสร้างแบบจำลอง (Model) ในการทำนายจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทย มีแนวโน้มจะมาประเทศไทยเพิ่มขึ้น ชุดข้อมูลที่ผู้วิจัยนำมาใช้ในงานวิจัย เป็นชุดข้อมูลของนักท่องเที่ยวชาวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2550 – 2558 ที่สำนักงานสถิติแห่งชาตินำออกมาเผยแพร่ผ่านเว็บไซต์ <http://www.nso.go.th>

สำหรับการวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพการพยากรณ์ จำนวนนักท่องเที่ยวชาวต่างประเทศที่เดินทางเข้ามาในประเทศไทย โดยผู้วิจัยเลือกชุดข้อมูลอนุกรมเวลาสำหรับการทดลองแบ่งเป็น 4 ชุดข้อมูล โดยเลือกประเทศที่มีจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทยมากที่สุด ได้แก่ ประเทศจีน ประเทศมาเลเซีย ประเทศญี่ปุ่น และประเทศเกาหลี ตามลำดับ

2. การเตรียมข้อมูล (Data Preparation)

งานวิจัยเวกเตอร์แมชชีนสำหรับการถดถอยนี้ เน้นการสร้างแบบจำลองเพื่อใช้ในการทำนายจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย โดยวิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) ได้แก่ การถดถอยเชิงเส้น (Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) โดยการนำแบบจำลองที่ได้มาเปรียบเทียบประสิทธิภาพ โดยแบ่งชุดข้อมูลออกเป็น 2 ส่วน คือ 1) ชุดข้อมูลการเรียนรู้ (Training Data Set) ใช้วิธีวัดรากที่สองของค่า

ความคลาดเคลื่อนที่กำลังสอง (Root Mean Square Error: RMSE) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เพื่อแสดงการเปรียบเทียบประสิทธิภาพแบบจำลอง เพื่อใช้ประมาณจำนวนนักท่องเที่ยวชาวต่างประเทศ 2) ชุดข้อมูลทดสอบ (Testing Data Set) ใช้วิธีการประมาณความแม่นยำในการพยากรณ์ด้วยค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error: MRE) สำหรับทดสอบประสิทธิภาพของแบบจำลองโดยแยกข้อมูลทดสอบรายปี และใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error: MMRE) เพื่อใช้ทดสอบประสิทธิภาพโดยรวมของแบบจำลองการพยากรณ์สำหรับชุดข้อมูลทดสอบ

3. การวิเคราะห์ข้อมูล

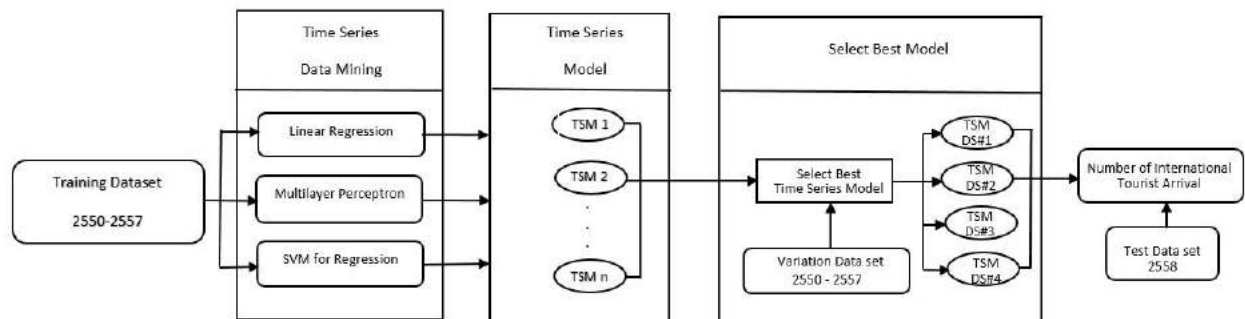
ผู้วิจัยเลือกใช้โปรแกรม WEKA Version 3.8.0 เพื่อทำการวิเคราะห์ข้อมูล และสร้างแบบจำลองการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย โดยวิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) ซึ่งวิธีการทั้งหมดจะได้ผลลัพธ์ในรูปแบบของโมเดล เป็นลักษณะของการแทนความรู้ (Knowledge Representation) แบบหนึ่ง รูปแบบการพยากรณ์เป็นการจัดเรียงข้อมูลแบบอนุกรมเวลา โดยสร้างแบบจำลองการพยากรณ์ โดยผู้วิจัยนำชุดข้อมูลในปี 2550 - 2557 เป็นชุดข้อมูลทดสอบ แบ่งเป็นชุดข้อมูลย้อนหลัง (Lagged) 1) 3 เดือน 2) 6 เดือน 3) 9 เดือน และ 4) 12 เดือน โดยผลลัพธ์ในการทดสอบประสิทธิภาพข้อมูลทั้ง 4 ชุด จะเป็นตัวชี้วัดอัลกอริทึมใด ที่มีความเหมาะสมกับการใช้ในการสร้างแบบจำลองให้กับประเทศใด สำหรับข้อมูลปี พ.ศ. 2558 นั้น เป็นค่าเป้าหมาย (Target) ในการพยากรณ์ประสิทธิภาพของการพยากรณ์ในแต่ละอัลกอริทึมจะเป็นดัชนีชี้วัดว่าอัลกอริทึมใดเหมาะสมกับการพยากรณ์ในเดือนใดในแต่ละไตรมาส รายละเอียดชุดข้อมูลตัวอย่าง นักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย ได้แก่ ประเทศจีน สามารถแสดงรายละเอียดได้จากตารางที่ 1

ตารางที่ 1 รายละเอียดการจัดเรียงข้อมูลแบบอนุกรมเวลาโดยข้อมูลมีลักษณะช่วง 12 เดือน

เดือน	จำนวนนักท่องเที่ยวประเทศจีนที่เดินทางเข้ามาในประเทศไทย								
	2550	2551	2552	2553	2554	2555	2556	2557	2558
	Training Data set / Variation Data set								Testing Data set
มกราคม	64,494	84,859	57,329	106,920	143,477	222,773	339,067	357,034	560,360
กุมภาพันธ์	92,237	113,990	53,106	152,717	183,163	158,742	464,975	360,023	780,516
มีนาคม	59,930	89,206	79,362	74,267	148,236	198,767	372,875	319,881	663,571
เมษายน	78,385	101,631	70,021	52,859	138,042	214,689	427,780	332,819	701,169
พฤษภาคม	66,692	81,564	36,425	35,833	134,825	192,642	383,292	302,823	668,079
มิถุนายน	62,521	61,970	35,370	49,110	124,605	180,481	370,064	220,497	632,889
กรกฎาคม	78,142	81,417	47,991	90,223	179,076	263,412	458,683	360,231	775,829
สิงหาคม	80,323	68,860	63,700	123,591	193,881	278,182	476,765	449,670	800,596
กันยายน	72,430	34,147	64,231	100,621	67,651	231,560	455,202	424,461	532,696
ตุลาคม	75,310	41,020	83,256	99,131	149,088	262,432	299,982	500,666	566,053
พฤศจิกายน	91,479	43,987	93,030	111,970	67,651	294,518	323,222	512,856	615,195
ธันวาคม	85,174	24,009	93,687	124,977	105,898	288,662	265,428	495,337	637,838
รวม	907,117	826,660	777,508	1,122,219	1,635,593	2,786,860	4,637,335	4,636,298	7,934,791

จากตารางที่ 1 แสดงรายละเอียดการจัดเรียงข้อมูลแบบอนุกรมเวลาโดยข้อมูลมีลักษณะช่วง 12 เดือน ผู้วิจัยจะนำข้อมูลอนุกรมเวลาทั้ง 3 เทคนิค ได้แก่ 1) Linear Regression 2) Multi-Layer Perceptron 3) Support Vector Machine for Regression เพื่อทำการทดสอบประสิทธิภาพของการทำนายจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย

ไทย โดยผู้วิจัยพิจารณาค่า RMSE และ MAE เป็นดัชนีชี้วัดแบบจำลองการทำนายที่มีความเหมาะสมกับการพยากรณ์ในแต่ละประเทศ สำหรับการทดสอบประสิทธิภาพของแบบจำลองโดยแยกเป็นแต่ละเดือนของชุดข้อมูลทดสอบ ผู้วิจัยจะพิจารณาจากค่า MRE และ MMRE เพื่อเลือกแบบจำลองที่เหมาะสม (Weiss and Indurkha, 1998) ในการนำไปใช้งาน



รูปที่ 2 ขั้นตอนการสร้างแบบจำลองเพื่อพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย

ผลการวิจัย

จากผลการทดลองได้ทำการวัดประสิทธิภาพแบบจำลอง เพื่อใช้ในการทำนายจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย โดยใช้วิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) ด้วยวิธีการดำเนินงานวิจัยโดยใช้ 3 เทคนิค กับชุดข้อมูลฝึกฝน (Training Data Set) และชุดข้อมูลทดสอบ (Test Data Set) จำนวน 4 ชุด ตามขั้นตอนวิธีวิจัย ผลการทดสอบสามารถแบ่งออกเป็น 2 ส่วนดังนี้

1. การเปรียบเทียบประสิทธิภาพของแบบจำลองทำนายข้อมูล จากชุดข้อมูลที่ถูกวิจัยเลือกใช้

2. จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทยมากที่สุด จำนวน 4 ประเทศ ซึ่งเป็นชุดข้อมูลสำหรับสร้างแบบจำลองการพยากรณ์ โดยผู้วิจัยจะนำข้อมูลในปี 2550– 2557 เป็นชุดข้อมูลย้อนหลัง (Lagged) 1) 3 เดือน 2) 6 เดือน 3) 9 เดือน และ 4) 12 เดือน นำมาสร้างแบบจำลองในการทำนายจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย เพื่อวิเคราะห์ประสิทธิภาพที่ได้จากเทคนิควิธีเหมืองข้อมูลทั้ง 3 แบบ โดยวิธีการเปรียบเทียบความคลาดเคลื่อนด้วยวิธี MAE และ RMSE จากผลการทดลองสามารถแสดงได้จากตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายโดยใช้ชุดข้อมูลย้อนหลัง (Lagged)

ชุดข้อมูล	Time Series Data Mining Techniques							
	ประเทศ	Quarter Lagged	Linear Regression		Multilayer Perception		SMOreg	
			MAE	RMSE	MAE	RMSE	MAE	RMSE
DS#1 จีน	Lag 3		37,739	56,320	36,979	54,208	37,809	56,754
	Lag 6		38,449	57,087	39,046	55,512	36,492	57,594
	Lag 9		33,723	50,099	30,297	39,367	31,506	51,699
	Lag 12		33,795	48,486	28,274	35,614	31,879	51,901
DS#2 มาเลเซีย	Lag 3		26,349	36,562	25,125	34,782	25,951	37,123
	Lag 6		26,764	37,024	24,280	34,027	25,361	37,194
	Lag 9		27,155	37,496	26,405	34,068	25,807	37,810
	Lag 12		25,517	32,264	19,549	25,831	23,530	32,208
DS#3 ญี่ปุ่น	Lag 3		12,166	15,972	13,254	17,030	12,123	16,185
	Lag 6		12,087	15,689	13,387	16,546	11,791	15,887
	Lag 9		12,076	15,719	10,074	12,826	11,090	15,905
	Lag 12		10,756	14,086	8,438	10,717	9,335	14,001
DS#4 เกาหลี	Lag 3		14,281	17,745	14,553	18,996	14,271	18,391
	Lag 6		13,689	16,590	11,456	14,899	13,297	17,156
	Lag 9		12,196	14,888	9,370	11,896	11,946	15,490
	Lag 12		11,429	14,116	7,534	10,086	11,044	14,827

จากตารางที่ 2 แสดงผลการทำนายของแต่ละเทคนิควิธีเหมืองข้อมูล กับการใช้ชุดข้อมูลโดยมีจำนวนวันย้อนหลัง (Lagged) ที่แตกต่างกันเมื่อใช้ค่า MAE และ RMSE เป็นดัชนีวัดประสิทธิภาพ จากการทดลองแสดงให้เห็นว่า ประเทศจีน (DS#1) เมื่อสร้างแบบจำลองด้วยเทคนิคการถดถอยเชิงเส้น ด้วยข้อมูลย้อนหลัง 12 เดือน มีประสิทธิภาพสูงที่สุด เนื่องจากค่าความแตกต่างของ MAE ข้อมูลย้อนหลัง 9 เดือน

(Lag 9) มีค่าความแตกต่างกันน้อยกว่า ค่าความแตกต่าง RMSE ข้อมูลย้อนหลัง 12 เดือน (Lag 12) แบบจำลองประสาทเทียมข้อมูลย้อนหลัง 12 เดือน (Lag 12) จึงมีประสิทธิภาพสูงที่สุด แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ด้วยข้อมูลย้อนหลัง 9 เดือน (Lag 9) ประเทศมาเลเซีย ประเทศญี่ปุ่น และประเทศเกาหลี พบว่าทั้ง 3 ประเทศใช้เทคนิคเหมืองข้อมูลทั้ง 3 แบบให้ค่าประสิทธิภาพสูงที่สุดเมื่อใช้ชุดข้อมูลย้อนหลัง 12 เดือน (Lag 12)

จากผลการทดลองผู้วิจัยจึงเลือกชุดข้อมูลย้อนหลังที่ทำให้ได้ค่าประสิทธิภาพที่ดีที่สุด ไปใช้ในการสร้างแบบจำลอง เพื่อทำนายจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทยในแต่ละเดือน

3. การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายในแต่ละเดือน จากการนำแบบจำลองการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดิน

ทางเข้ามาในประเทศไทยในปี 2558 โดยแยกปริมาณออกเป็นแต่ละเดือนจำนวน 12 เดือน ด้วยวิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) โดยใช้อัลกอริทึม 3 แบบ และคำนวณหาค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error) สามารถแสดงผลได้จากตารางที่ 3

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายในแต่ละเดือน

ชุดข้อมูล		Times Series Data Mining Techniques						
ประเทศ	ข้อมูล 2558		Linear Regression		Multilayer Perception		SMOreg	
	Month	Actual	Estimate	MRE	Estimate	MRE	Estimate	MRE
DS#1 จีน	มกราคม	560,360	550,143.89	0.018	1,014,309.03	0.810	594,817.30	0.061
	กุมภาพันธ์	780,516	622,098.83	0.203	631,030.51	0.192	612,935.31	0.215
	มีนาคม	663,571	603,882.88	0.090	812,607.25	0.225	607,883.76	0.084
	เมษายน	701,169	605,715.69	0.136	565,078.19	0.194	639,260.61	0.088
	พฤษภาคม	668,079	592,065.64	0.114	339,798.24	0.491	584,670.21	0.125
	มิถุนายน	632,889	555,477.38	0.122	327,779.45	0.482	566,326.32	0.105
	กรกฎาคม	775,829	572,818.32	0.262	697,637.90	0.101	610,517.78	0.213
	สิงหาคม	800,596	595,944.41	0.256	403,072.19	0.497	600,292.46	0.250
	กันยายน	532,696	601,460.44	0.129	688,732.19	0.293	610,051.40	0.145
	ตุลาคม	566,053	594,651.08	0.051	820,545.40	0.450	630,466.08	0.114
	พฤศจิกายน	615,195	590,793.06	0.040	218,539.67	0.645	599,757.10	0.025
	ธันวาคม	637,838	582,311.93	0.087	218,539.67	0.657	623,794.32	0.022
	MMRE			12.56%		41.96%		12.07%
DS#2 มาเลเซีย	มกราคม	259,863	260,708.20	0.003	143,807.39	0.447	268,458.24	0.033
	กุมภาพันธ์	275,563	266,584.35	0.033	200,764.67	0.271	284,756.41	0.033
	มีนาคม	278,509	277,215.83	0.005	251,552.65	0.097	297,110.74	0.067
	เมษายน	271,568	251,781.26	0.073	329,859.53	0.215	280,825.83	0.034
	พฤษภาคม	298,671	245,116.26	0.179	285,158.76	0.045	278,907.79	0.066
	มิถุนายน	290,790	226,642.75	0.221	263,726.47	0.093	259,186.04	0.109
	กรกฎาคม	288,133	227,985.52	0.209	75,111.12	0.739	256,158.34	0.111
	สิงหาคม	282,622	209,196.50	0.260	27,894.48	0.901	260,153.17	0.080
	กันยายน	260,029	192,727.91	0.259	192,016.54	1.738	240,159.27	0.076
	ตุลาคม	249,171	222,621.48	0.107	273,798.38	2.099	248,647.15	0.002
	พฤศจิกายน	293,217	229,627.49	0.217	349,505.49	2.192	243,054.75	0.171
	ธันวาคม	375,261	277,679.23	0.260	320,218.56	1.853	289,961.11	0.227
	MMRE			15.20%		89.09%		8.41%

ชุดข้อมูล		Times Series Data Mining Techniques						
ประเทศ	ข้อมูล 2558		Linear Regression		Multilayer Perception		SMOreg	
	Month	Actual	Estimate	MRE	Estimate	MRE	Estimate	MRE
DS#3 ญี่ปุ่น	มกราคม	121,825	126,609.69	0.039	112,597.95	0.076	128,953.20	0.059
	กุมภาพันธ์	129,344	117,740.02	0.090	99,424.13	0.231	123,193.51	0.048
	มีนาคม	118,198	118,536.02	0.003	93,202.61	0.211	119,500.36	0.011
	เมษายน	102,218	116,616.59	0.141	95,803.00	0.063	113,059.34	0.106
	พฤษภาคม	103,601	104,605.72	0.010	87,406.34	0.156	101,825.06	0.017
	มิถุนายน	100,572	97,015.02	0.035	78,934.86	0.215	96,312.52	0.042
	กรกฎาคม	111,394	99,932.90	0.103	74,595.81	0.330	103,468.93	0.071
	สิงหาคม	141,672	110,473.03	0.220	80,552.94	0.431	105,567.95	0.255
	กันยายน	115,137	104,103.02	0.096	59,523.02	0.483	96,143.05	0.165
	ตุลาคม	97,043	98,066.30	0.011	58,712.63	0.395	89,942.78	0.073
	พฤศจิกายน	115,525	101,946.07	0.118	65,369.62	0.434	94,438.99	0.183
	ธันวาคม	125,161	110,757.81	0.115	86,098.24	0.312	102,755.11	0.179
	MMRE		8.17%		27.82%		10.07%	
DS#4 เกาหลี	มกราคม	153,062	132,030.21	0.137	148,998.84	0.027	128,444.60	0.161
	กุมภาพันธ์	117,331	123,193.70	0.050	90,650.25	0.227	128,239.09	0.093
	มีนาคม	100,236	106,682.96	0.064	90,753.70	0.095	103,331.94	0.031
	เมษายน	86,307	85,164.92	0.013	78,806.82	0.087	86,888.11	0.007
	พฤษภาคม	89,584	93,711.96	0.046	76,699.54	0.144	92,731.65	0.035
	มิถุนายน	91,776	91,035.72	0.008	112,962.71	0.231	89,901.57	0.020
	กรกฎาคม	121,499	99,308.71	0.183	149,502.32	0.230	105,486.43	0.132
	สิงหาคม	137,405	98,394.24	0.284	86,236.07	0.372	105,679.08	0.231
	กันยายน	94,249	82,235.99	0.127	116,399.17	0.235	88,209.31	0.064
	ตุลาคม	108,923	84,051.57	0.228	134,731.33	0.237	83,404.37	0.234
	พฤศจิกายน	123,303	86,203.00	0.301	112,689.74	0.086	81,578.72	0.338
	ธันวาคม	149,320	101,313.18	0.322	203,181.82	0.361	94,271.48	0.369
	MMRE		14.70%		19.43%		14.29%	

เมื่อพิจารณาผลการทดลองในตารางที่ 3 พบว่าการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย นำมาทดสอบกับชุดข้อมูลจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทยในปี พ.ศ. 2558 การสร้างแบบจำลองด้วยเทคนิคการถดถอยเชิงเส้น ให้ค่าประสิทธิภาพที่ดีที่สุด ได้แก่ ประเทศญี่ปุ่น มีค่า MMRE เท่ากับ 8.17% การสร้างแบบจำลองด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยให้ค่าประสิทธิภาพสูงสุด 3 ประเทศ ได้แก่ ประเทศจีน มีค่า MMRE เท่ากับ 12.07% ประเทศมาเลเซีย มีค่า MMRE เท่ากับ 8.41%

และประเทศเกาหลี มีค่า MMRE เท่ากับ 14.29%

วิจารณ์

งานวิจัยนี้เป็นการสร้างแบบจำลองการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย โดยใช้วิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) โดยใช้เทคนิคเหมืองข้อมูล 3 เทคนิคคือ การถดถอยเชิงเส้น (Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support

Vector Machine for Regression) โดยมีขั้นตอนการวิจัยดังนี้ 1) ศึกษาปัญหาและวิเคราะห์ข้อมูล จากการศึกษาข้อมูลที่สำนักงานสถิติแห่งชาตินำมาเผยแพร่ผ่านทางเว็บไซต์ ในปี พ.ศ. 2550 – 2558 มาใช้ในการทดลอง 2) การเตรียมข้อมูล (Data Preparation) ผู้วิจัยได้เลือกข้อมูล 4 ประเทศที่นักท่องเที่ยวเดินทางเข้ามาในประเทศไทยมากที่สุด ได้แก่ ประเทศจีน ประเทศมาเลเซีย ประเทศญี่ปุ่น และประเทศเกาหลี โดยแบ่งเป็นชุดข้อมูลรายเดือน ตั้งแต่ปี พ.ศ. 2550 – 2558 จำนวน 9 ปี หรือ 108 เดือน 3) การวิเคราะห์ข้อมูล (Data Analysis) รูปแบบการพยากรณ์มีลักษณะเป็นการจัดเรียงข้อมูลแบบอนุกรมเวลา โดยการสร้างชุดข้อมูลสำหรับการสร้างแบบจำลองการพยากรณ์ ผู้วิจัยนำข้อมูลในปี พ.ศ. 2550 – 2557 มาเป็นชุดข้อมูลทดสอบ โดยแบ่งเป็น ชุดข้อมูลย้อนหลัง (Lagged) 1) 3 เดือน 2) 6 เดือน 3) 9 เดือน และ 4) 12 เดือน

ผลการวิจัยพบว่า การวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมือนข้อมูลนั้น การสร้างแบบจำลองการพยากรณ์ด้วยวิธี โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น ให้ค่าความผิดพลาด 27.82% เมื่อวัดด้วยค่า MMRE โดยค่าความผิดพลาดจะลดลงเมื่อใช้วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยค่า MMRE เท่ากับ 10.07% และค่าความผิดพลาดน้อยที่สุดด้วยเทคนิคการถดถอยเชิงเส้น โดยค่า MMRE เท่ากับ 8.17% จึงเป็นวิธีที่มีประสิทธิภาพมากที่สุดที่ใช้วิเคราะห์ข้อมูลประเทศญี่ปุ่น การเลือกใช้เทคนิคโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น ให้ค่าความผิดพลาด 41.96% เมื่อวัดด้วยค่า MMRE ค่าความผิดพลาดจะลดลงเมื่อใช้วิธีการถดถอยเชิงเส้น โดยค่า MMRE เท่ากับ 12.56% และค่าความผิดพลาดน้อยที่สุดด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยค่า MMRE เท่ากับ 12.07% จึงเป็นวิธีที่มีประสิทธิภาพมากที่สุดที่ใช้วิเคราะห์ข้อมูลประเทศจีน การเลือกใช้เทคนิคโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น ให้ค่าความผิดพลาด 89.09% เมื่อวัดด้วยค่า MMRE ค่าความผิดพลาดจะลดลงเมื่อใช้วิธีการ

ถดถอยเชิงเส้น โดยค่า MMRE เท่ากับ 15.20% และค่าความผิดพลาดน้อยที่สุดด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยค่า MMRE เท่ากับ 8.41% จึงเป็นวิธีที่มีประสิทธิภาพมากที่สุดที่ใช้วิเคราะห์ข้อมูลประเทศมาเลเซีย การเลือกใช้เทคนิคโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น ให้ค่าความผิดพลาด 19.43% เมื่อวัดด้วยค่า MMRE ค่าความผิดพลาดจะลดลงเมื่อใช้วิธีการถดถอยเชิงเส้น โดยค่า MMRE เท่ากับ 14.70% และค่าความผิดพลาดน้อยที่สุดด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยค่า MMRE เท่ากับ 14.29% จึงเป็นวิธีที่มีประสิทธิภาพมากที่สุดที่ใช้วิเคราะห์ข้อมูลประเทศเกาหลี ดังนั้นการสร้างแบบจำลองด้วยเทคนิคการถดถอยเชิงเส้น มีความเหมาะสมกับประเทศญี่ปุ่น ด้วยค่า MMRE เท่ากับ 8.17% การสร้างแบบจำลองด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีความเหมาะสมกับ 3 ประเทศ คือ ประเทศจีน มีค่า MMRE เท่ากับ 12.07% ประเทศมาเลเซีย มีค่า MMRE เท่ากับ 8.41% และประเทศเกาหลี มีค่า MMRE เท่ากับ 14.29%

เมื่อพิจารณาผลการวิจัยอย่างละเอียดพบว่า ประเด็นหลักในการพยากรณ์จำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย แบ่งเป็น 3 ประเด็น ได้แก่ 1) ผู้พยากรณ์ต้องเลือกเทคนิคที่มีความเหมาะสมกับชุดของข้อมูล เมื่อผู้วิจัยพิจารณาแล้ว พบว่าเทคนิคที่แตกต่างกัน จะมีผลต่อประสิทธิภาพของการทำงานที่ต่างกัน และจำนวนข้อมูลย้อนหลัง (Lagged) มีผลต่อประสิทธิภาพของแบบจำลองด้วย 2) การเลือกเทคนิควิธีที่ใช้กับข้อมูลทดสอบ (Testing Data Set) ซึ่งเป็นตัวแทนในการทำนาย ควรเลือกชุดข้อมูลสุดท้ายที่มีความครบถ้วน ในที่นี้ผู้วิจัยเลือกปี พ.ศ. 2558 เป็นชุดข้อมูลทดสอบ ซึ่งจำนวนชุดของข้อมูลครบถ้วนทั้งหมด 12 เดือน ซึ่งเป็นชุดข้อมูลที่สามารถใช้เพื่อการทดสอบ 3) จากผลการทดลองแสดงให้เห็นรายละเอียดในแต่ละเดือน โดย การวิจัยแบ่งข้อมูลออกเป็น 12 เดือน พบว่าประสิทธิภาพของ MAE ในแต่ละเดือนที่แตกต่างกัน

ดังนั้นการพิจารณาจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทยนั้น มีปริมาณที่แตกต่างกันไปในแต่ละเดือน ซึ่งจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในบางเดือนอาจมีจำนวนน้อยมาก ทั้งนี้อาจขึ้นอยู่กับปัญหาภายในประเทศของนักท่องเที่ยว เช่น ปัญหาด้านเศรษฐกิจ การเมือง และปัญหาด้านก่อการร้าย ซึ่งมีผลต่อการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) เนื่องจากมีผลต่อประสิทธิภาพการวัดด้วยค่าความคาดเคลื่อนสัมพัทธ์ (Magnitude of the Relative Error: MRE) ซึ่งส่งผลกระทบต่อประสิทธิภาพของค่าเฉลี่ยความคาดเคลื่อนสัมพัทธ์ (Mean Magnitude of the Relative Error: MMRE) ที่ใช้ในชุดข้อมูลทดลอง ในการหลีกเลี่ยงค่าความเอนเอียง (Bias) นั้นไม่ควรเลือกการตัดข้อมูลที่มีค่าน้อยเกินไปหรือมากเกินไปในชุดข้อมูลทั้ง 3 เทคนิค แม้ว่าจำนวนนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทยในแต่ละเดือนนั้นมีความแตกต่างกันมากทำให้ค่า MMRE มีความคลาดเคลื่อนสูง เพื่อให้ผลของการพยากรณ์และวิเคราะห์ข้อมูลที่ต้องการ และตรงตามความเป็นจริงที่ใช้ในการพยากรณ์มากที่สุด

คำขอบคุณ

ผู้วิจัยขอขอบคุณ ผศ.ดร.วีระยุทธ พิมพาภรณ์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา ผู้ให้คำแนะนำในการทำวิจัยในครั้งนี้ และแหล่งข้อมูลต่าง ๆ เครื่องมือที่ใช้ในการทำวิจัย จากมหาวิทยาลัยสยาม

เอกสารอ้างอิง

- Abran, A. 2015. Software Project Estimation: The Fundamentals for Providing High Quality Information to Decision Makers. John Wiley & Sons, New Jersey, 288 p.
- Elfeky, Mohamed G, Aref, Walid G., Elmagarmid, Ahmed K. (2005). Periodicity Detection in Time Series Database, pp. 1-13. Journal IEEE Transactions on Knowledge and Data Engineering. Vol. 17 No. 7, July 2005.
- Han, J., Pei, J., & Kamber, M. 2011. Data Mining: Concepts and Techniques. 2nd ed. Morgan Kaufmann, San Francisco, 745 p.
- Malode, Yogesh, Patel, Rahila. Review on Periodicity Mining Techniques in Time Series Data, pp. 217- 222. International Journal of Advanced Computer Research Vol. 2 No. 4 Issue 6 Dec 2012.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. 2015. Introduction to Linear Regression Analysis. 5th ed., John Wiley & Sons Inc. New Jersey, 672 p.
- Neter, John. 1989. Applied Linear Regression Models. Irwin, Boston, 547 p.
- Seber, G. A., & Lee, A. J. 2003. Linear Regression Analysis. 2nd ed., John Wiley & Sons, New Zealand, 549 p.
- Villordon, A., Clark, C., Ferrin, D., & LaBonte, D. 2009. Using Growing Degree Days, Agrometeorological Variables, Linear Regression, and Data Mining Methods to Help Improve Prediction of Sweet potato Harvest, pp. 133-144. In Louisiana. HortTechnology, Vol. 19 No. 1 Jan - Mar, 2009.
- Vinod, V., & Leela, S. 2013. A Survey on Periodicity Detection Techniques in Time Series Databases. In SRSA Publications. The International Journal of Engineering Research and Technology, Vol.2 Issue 12 Dec 2013.
- Witten, I. H., & Frank, E. 2005. Data Mining: Practical Machine Learning Tools and Techniques. 2nd ed., Morgan Kaufmann, San Francisco, 558 p.