การเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์อัตราการว่างงานในประเทศไทย พ.ศ. 2564 - 2567

A Comparative Analysis of Models for Forecasting Thailand's Unemployment Rate, 2021-2024

*สุรีนาฎ มะโนลา**
*Sureenat Manola**
*ธีรภพ แสงศรี***
*Theeraphop Saengsri***
*เทวา พรหมนุชานนท์****
*Tewa Promnuchanon****

รับบทความ: 30 ตุลาคม 2568 / แก้ไขบทความ: 12 ธันวาคม 2568/ ตอบรับการตีพิมพ์: 23 ธันวาคม 2568

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาผลการเปรียบเทียบแบบจำลองการพยากรณ์อัตราการว่างงาน ได้แก่ แบบจำลองอารีมา (ARIMA) แบบจำลองโฮลต์–วินเทอร์ส (Holt's Winters) แบบจำลองเคเนียเรสต์เนเบอร์ส (K-NN) และการถดถอยเชิงเส้น (Linear Regression) โดยใช้ชุดข้อมูลอัตราการว่างงานของประเทศไทยจากกองบริหารข้อมูลตลาดแรงงาน ซึ่งเป็นข้อมูลรายจังหวัดจำนวน 77 จังหวัด ครอบคลุมช่วงปี พ.ศ. 2564–2567 เพื่อนำมาประเมินประสิทธิภาพของแบบจำลองและเปรียบเทียบความแม่นยำในการพยากรณ์ ผลการวิจัยพบว่าแบบจำลองอารีมา (ARIMA) มีค่า MAE (Mean Absolute Error) ต่ำที่สุดเท่ากับ 6,658.515 และค่า RMSE (Root Mean Square Error) ต่ำที่สุดเท่ากับ 8,578.801 สะท้อนถึงความคลาดเคลื่อนเชิงตัวเลขที่น้อยและมีเสถียรภาพสูง แม้ว่าแบบจำลองโฮลต์–วินเทอร์ส (Holt's Winters) จะมีค่า MAPE (Mean Absolute Percentage Error) ต่ำที่สุดเท่ากับร้อยละ 4 ซึ่งแสดงถึงความแม่นยำเชิงสัมพัทธ์ที่ดี แต่ยังมีค่า RMSE และ MAE สูงกว่าแบบจำลองอารีมา (ARIMA) ขณะที่แบบจำลองเคเนียเรสต์เนเบอร์ส (K-NN) มีค่า MAPE อยู่ในระดับปานกลางที่ร้อยละ 6 แต่มีค่า RMSE สูงมากเท่ากับ 62,036.073 แสดงถึงความไม่เสถียรของการพยากรณ์ ส่วนการถดถอยเชิงเส้น (Linear Regression) แม้จะมีค่า MAPE เท่ากับร้อยละ 0 แต่กลับมีค่า RMSE และ MAE สูงที่สุดอย่างผิดปกติ สะท้อนถึงปัญหาการเรียนรู้เกินชุดข้อมูล (Overfitting) และไม่เหมาะสมต่อการใช้งานจริง ดังนั้น แบบจำลองอารีมา (ARIMA) จึงเป็นแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์อัตราการว่างงานของประเทศไทยในงานวิจัยนี้

คำสำคัญ: แบบจำลอง การพยากรณ์ข้อมูล อัตราการว่างงาน

---

*_**_*** หลักสูตรระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา

*_**_*** Department of Business Information System, Faculty of Business Administration and Liberal Arts, Rajamamgala University of Technology Lanna

** Corresponding author E-mail address : tees@rmutl.ac.th

Abstract

This study aims to compare the performance of unemployment rate forecasting models, including the Autoregressive Integrated Moving Average (ARIMA), Holt's Winters, K-Nearest Neighbors (K-NN), and Linear Regression models. The analysis is based on Thailand's unemployment rate data obtained from the Department of Employment, consisting of provincial-level data from 77 provinces covering the period from 2021 to 2024. The objective is to evaluate the performance of each model and compare their forecasting accuracy. The results indicate that the ARIMA model yields the lowest Mean Absolute Error (MAE) of 6,658.515 and the lowest Root Mean Square Error (RMSE) of 8,578.801, reflecting lower numerical forecasting errors and greater stability. Although Holt's Winters model achieves the lowest Mean Absolute Percentage Error (MAPE) at 4%, indicating high relative accuracy, its RMSE and MAE values remain higher than those of the ARIMA model. The K-Nearest Neighbors (K-NN) model shows a moderate MAPE of 6% but exhibits a very high RMSE of 62,036.073, suggesting instability in its forecasting results. Meanwhile, the Linear Regression model records a MAPE of 0%, but its RMSE and MAE values are abnormally high, indicating an overfitting problem and poor suitability for practical applications. Therefore, the findings conclude that the ARIMA model is the most appropriate approach for forecasting Thailand's unemployment rate, as it provides the best balance between forecasting accuracy and numerical error.

Keywords: Models, Data forecasting, Unemployment rate

บทนำ

อัตราการว่างงานเป็นตัวชี้วัดที่สำคัญในการประเมินภาวะเศรษฐกิจและสังคมของประเทศ โดยสะท้อนถึงความสามารถของตลาดแรงงานในการรองรับแรงงานที่ต้องการมีงานทำ การเปลี่ยนแปลงของอัตราการว่างงานจึงสามารถบ่งชี้ถึงเสถียรภาพทางเศรษฐกิจ ความมั่นคงของรายได้ประชากร ตลอดจนคุณภาพชีวิตโดยรวมของประชาชน ทั้งนี้ หากอัตราการว่างงานอยู่ในระดับสูงอย่างต่อเนื่อง อาจส่งผลให้เกิดปัญหาทางเศรษฐกิจและสังคมหลายประการ เช่น การลดลงของรายได้ครัวเรือน การเพิ่มขึ้นของความยากจน และปัญหาด้านสุขภาพจิตของประชาชน (กองบริหารตลาดแรงงาน, 2567)

ในช่วงปี พ.ศ. 2564–2567 ประเทศไทยเผชิญกับความเปลี่ยนแปลงทางเศรษฐกิจและสังคมอย่างมีนัยสำคัญจากหลายปัจจัย ไม่ว่าจะเป็นการแพร่ระบาดของโรคโควิด-19 การปรับเปลี่ยนนโยบายเศรษฐกิจและการพัฒนาเทคโนโลยีที่ส่งผลต่อโครงสร้างตลาดแรงงาน ทำให้อัตราการว่างงานเกิดความผันผวนอย่างชัดเจน การวิเคราะห์แนวโน้มและรูปแบบของอัตราการว่างงานในช่วงเวลาดังกล่าวจึงมีความจำเป็น เพื่อให้สามารถเข้าใจสถานการณ์ได้อย่างลึกซึ้งและนำไปสู่การกำหนดนโยบายด้านแรงงานและเศรษฐกิจที่เหมาะสม

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่างานวิจัยทั้งในและต่างประเทศได้นำแบบจำลองอนุกรมเวลาและแบบจำลองเชิงสถิติมาใช้ในการพยากรณ์อัตราการว่างงานอย่างแพร่หลาย โดยแบบจำลองอารีมา

(ARIMA) และแบบจำลองโฮลต์–วินเทอร์ส (Holt's Winters) มักถูกนำมาใช้กับข้อมูลที่มีลักษณะแนวโน้มและฤดูกาล ขณะที่แบบจำลองการเรียนรู้ของเครื่อง เช่น เคเนียเรสต์เนเบอร์ส (K-Nearest Neighbors: K-NN) และการถดถอยเชิงเส้น (Linear Regression) ถูกนำมาใช้เพื่อศึกษาความสัมพันธ์ของข้อมูลและเพิ่มความยืดหยุ่นในการพยากรณ์ อย่างไรก็ตาม งานวิจัยส่วนใหญ่ยังมุ่งเน้นการใช้แบบจำลองเพียงบางประเภท หรือประเมินประสิทธิภาพของแบบจำลองโดยใช้ตัวชี้วัดที่แตกต่างกัน ทำให้การเปรียบเทียบผลลัพธ์ระหว่างแบบจำลองยังไม่ชัดเจน โดยเฉพาะในบริบทของประเทศไทยและในช่วงหลังการแพร่ระบาดของโรคโควิด-19 ซึ่งยังมีงานศึกษาที่เปรียบเทียบแบบจำลองเชิงสถิติและแบบจำลองการเรียนรู้ของเครื่องร่วมกันอย่างเป็นระบบค่อนข้างจำกัด (Box et al., 2016; Hyndman & Athanasopoulos, 2021; Makridakis et al., 2018)

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบและประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์อัตราการว่างงานของประเทศไทยในช่วงปี พ.ศ. 2564–2567 โดยใช้ข้อมูลจากกรมการจัดหางาน กองบริหารข้อมูลตลาดแรงงาน สำนักงานสถิติแห่งชาติ (2566) มาวิเคราะห์ด้วยเทคนิคการวิเคราะห์อนุกรมเวลา (Time Series Analysis) ผ่านการเปรียบเทียบแบบจำลองการพยากรณ์ ได้แก่ แบบจำลองอารีมา (ARIMA) แบบจำลองโฮลต์–วินเทอร์ส (Holt's Winters) แบบจำลองเคเนียเรสต์เนเบอร์ส (K-NN) และการถดถอยเชิงเส้น (Linear Regression) ภายใต้กระบวนการ CRISP-DM โดยมีการประเมินประสิทธิภาพของแบบจำลองด้วยตัวชี้วัดความแม่นยำ ได้แก่ ค่าร้อยละความคลาดเคลื่อนเฉลี่ย (MAPE) ค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (RMSE) และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAE) เพื่อให้ได้ข้อสรุปเชิงเปรียบเทียบถึงความเหมาะสมของแบบจำลองแต่ละประเภท นอกจากนี้ ผลการวิเคราะห์ยังถูกนำเสนอในรูปแบบการแสดงผลข้อมูล (Data Visualization) ด้วยโปรแกรม Power BI

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองเพื่อการพยากรณ์อัตราการว่างงานในประเทศไทยในช่วงปี พ.ศ. 2564 - 2567
2. เพื่อประเมินประสิทธิภาพของแบบจำลองพยากรณ์อัตราการว่างงานในประเทศไทยในช่วงปี พ.ศ.2564 - 2567

วิธีดำเนินการวิจัย

การวิเคราะห์ในรูปแบบของการทำเหมืองข้อมูลโดยใช้กระบวนการวิเคราะห์ข้อมูลด้วย Cross Industry Standard Process for Data Mining หรือ CRISP-DM (ฐาปณีย์ บุญชอบ, 2563) มีขั้นตอนดังนี้

1. กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding) วัตถุประสงค์ของการวิจัยนี้คือการพยากรณ์อัตราการว่างงานของประเทศไทยในช่วงปี พ.ศ. 2564–2567 โดยใช้ข้อมูลอัตราการว่างงานในอดีตมาวิเคราะห์และสร้างแบบจำลองทางสถิติที่เหมาะสม เพื่อให้สามารถคาดการณ์แนวโน้มของอัตราการว่างงานในอนาคตได้อย่างแม่นยำ ซึ่งผลลัพธ์จากการวิจัยจะนำไปใช้เป็นข้อมูลประกอบการตัดสินใจของหน่วยงานที่เกี่ยวข้อง ทั้งในการกำหนดนโยบายด้านแรงงาน และการวางแผนทางเศรษฐกิจของภาครัฐอย่างมีประสิทธิภาพ

2. การทำความเข้าใจข้อมูล (Data Understanding) ทำการรวบรวมข้อมูลเพื่อตรวจสอบรายละเอียดและปริมาณ จากเว็บไซต์ www.doe.go.th ซึ่งเป็นเว็บไซต์ของกองบริหารข้อมูลตลาดแรงงาน สามารถเชื่อถือได้ เป็นช่องทางให้ ผู้ใช้บริการทั้งภาคประชาชน ภาคธุรกิจเอกชน รวมถึงหน่วยงานของรัฐ สามารถค้นหา และ เข้าถึงข้อมูลที่มีคุณภาพของภาครัฐได้ง่าย

3. การเตรียมข้อมูล (Data Preparation) นำชุดข้อมูลมาทำการคัดเลือก ข้อมูล และทำการ Data Cleaning แก้ไขข้อมูลที่ผิดพลาด ทำการจัดหมวดหมู่เพื่อความถูกต้อง ข้อมูลดำเนินการจัดกลุ่มแบ่งข้อมูล 3 ตัวแปร (Attribute) คือ ผู้มีงานทำ ผู้ว่างงาน และผู้ที่รอฤดูกาลเพื่อแบ่งข้อมูลอย่างชัดเจนเพื่อให้ได้ข้อมูลที่สมบูรณ์

3.1 การทำความสะอาดข้อมูล (Data Cleaning) ตรวจสอบ แก้ไขหรือลบรายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ได้ดำเนินการกำจัดแถวรายการที่มีข้อมูลไม่ถูกต้องข้อมูลที่ผิดพลาด สูญหาย ไม่ถูกต้อง หรือยังไม่สมบูรณ์ได้แก่ ค่าว่างและ n.a. จำนวน 284 ค่า สำหรับค่าว่าง (Missing Values) และค่า n.a. ที่พบในชุดข้อมูล จำนวน 284 ค่า ได้ดำเนินการจัดการโดยการ แทนค่าทั้งหมดด้วยค่า null เนื่องจากในตัวแปร (Attribute) ที่มีค่าที่หายไปคือผู้ที่รอฤดูกาลซึ่งหมายถึงไม่มีการจ้างงานเกิดขึ้น เพื่อให้สามารถระบุและจัดการข้อมูลที่ไม่สมบูรณ์ได้อย่างชัดเจนในการวิเคราะห์ต่อไป ข้อมูลที่เหลือหลังจากการแทนค่า null ได้รับการตรวจสอบความถูกต้องและความสมบูรณ์

3.2 การจัดหมวดหมู่ (Transform) นำชุดข้อมูลผ่านการคัดเลือกข้อมูลและการทำ Data Cleaning มาทำการจัดหมวดหมู่เพื่อความถูกต้องและความชัดเจนของข้อมูล ผู้วิเคราะห์ข้อมูลดำเนินการจัดกลุ่มแบ่งข้อมูลออกเป็น 3 ตัวแปร (Attribute) ได้แก่ ผู้มีงานทำ (Have Work) ผู้ว่างงาน (Unemployee) และผู้ที่รอฤดูกาล (Waiting for Work) เนื่องจากตัวแปร (Attribute) ทั้งสามสามารถสะท้อนสถานะแรงงานได้อย่างชัดเจนและครอบคลุมทุกกลุ่มหลักของตลาดแรงงาน การแบ่งกลุ่มเช่นนี้ช่วยให้การวิเคราะห์ข้อมูลเชิงสถิติและการสร้างแบบจำลองพยากรณ์สามารถทำได้อย่างแม่นยำ

3.3 การโหลดข้อมูล (Loading) เข้าสู่โปรแกรม Rapid miner Studio เพื่อจะนำไปวิเคราะห์และสร้างออกมาเป็นรายงานภาพ Visualization

4. การสร้างแบบจำลอง (Modeling) ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลด้วยกระบวนการทำเหมืองข้อมูล โดยใช้เทคนิคการวิเคราะห์อนุกรมเวลา (Time Series Analysis) และเลือกใช้แบบจำลองในการพยากรณ์จำนวน 4 แบบ ได้แก่ แบบจำลองอารีมา (ARIMA) แบบจำลองโฮลต์–วินเทอร์ส (Holt–Winters) แบบจำลองเคเนียเรสต์เนเบอร์ส (K-NN) และการถดถอยเชิงเส้น (Linear Regression) ผู้วิจัยเลือกใช้แบบจำลองทั้ง 4 แบบ เพื่อให้ครอบคลุมทั้งแบบจำลองเชิงสถิติและแบบจำลองการเรียนรู้ของเครื่อง ซึ่งมีหลักการทำงานและสมมติฐานที่แตกต่างกัน การเปรียบเทียบแบบจำลองดังกล่าวช่วยให้สามารถประเมินความเหมาะสมของแนวทางแต่ละประเภทในการพยากรณ์อัตราการว่างงานได้อย่างเป็นระบบ การวิเคราะห์ข้อมูลดำเนินการด้วยโปรแกรม RapidMiner Studio โดยใช้ชุดข้อมูลจำนวน 8 ตัวแปร (Attributes) ซึ่งตัวแปรหลักที่ใช้ในการพยากรณ์ ได้แก่ จำนวนผู้ว่างงาน

4.1 แบบจำลองอารีมา (ARIMA) เป็นแบบจำลองเชิงสถิติที่นิยมใช้ในการพยากรณ์ข้อมูลอนุกรมเวลา โดยอาศัยความสัมพันธ์ของข้อมูลในอดีต (Box et al., 2016) ที่ได้รับความนิยมอย่างแพร่หลาย โดยผสมผสานระหว่างการถดถอยอัตโนมัติ (AR) การทำให้ข้อมูลอยู่ในระดับคงที่ (I) และค่าเฉลี่ยเคลื่อนที่ (MA) เพื่อสร้างแบบจำลองที่สามารถพยากรณ์ข้อมูลในอนาคตได้อย่างมีประสิทธิภาพ ผู้วิเคราะห์ได้นำเอา

แบบจำลอง ARIMA มาใช้ในการวิเคราะห์ข้อมูลอัตราการว่างงานในประเทศไทยปี 2564-2567 โดยใช้โปรแกรม Rapid miner studio ในการวิเคราะห์ข้อมูลส่วนประกอบของสมการ ARIMA AR (Autoregressive) ส่วนนี้แสดงถึงการพึ่งพาของค่าปัจจุบันกับค่าที่ผ่านมา (Lagged values) ซึ่งสามารถเขียนได้ในรูปสมการ

$$Yt = \phi 1Yt - 1 + \phi 2Yt - 2 + \cdots + \phi pYt - p + \epsilon t$$

โดยที่

Yt ค่าของตัวแปร ณ เวลา t

ϕi ค่าสัมประสิทธิ์ AR

p จำนวน lag

ϵt ค่า error หรือ residual

I (Integrated): ส่วนนี้ใช้เพื่อทำให้ข้อมูลอนุกรมเวลามีความนิ่ง (Stationary) โดยการนำค่าความต่าง (Difference) มาใช้

$$Yt' = Yt - Yt - 1$$

โดยที่

Yt' ค่าความต่างอันดับที่ 1 (First-order difference)

MA (Moving Average) ส่วนนี้แสดงถึงการพึ่งพาของค่าปัจจุบันกับค่า residual จาก lag ที่ผ่านมา

$$Yt = \epsilon t + \theta 1\epsilon t - 1 + \theta 2\epsilon t - 2 + \cdots + \theta q\epsilon t - q$$

โดยที่

ϵt ค่า error ณ เวลา t

θi ค่าสัมประสิทธิ์ MA

q จำนวน lag ของ residual

สมการ ARIMA ที่รวมทุกส่วนเข้าด้วยกัน

$$\Phi p(B)(1 - B)dYt = \Theta q(B)\epsilon t$$

โดยที่

$\Phi p(B) = 1 - \phi 1B - \phi 2B2 - \cdots - \phi pBp$ ส่วน AR

$\Theta q(B) = 1 + \theta 1B + \theta 2B2 + \cdots + \theta qBq$ ส่วน MA

$(1 - B)d$ การทำ Differencing

B Backshift operator $(BYt = Yt - 1)$

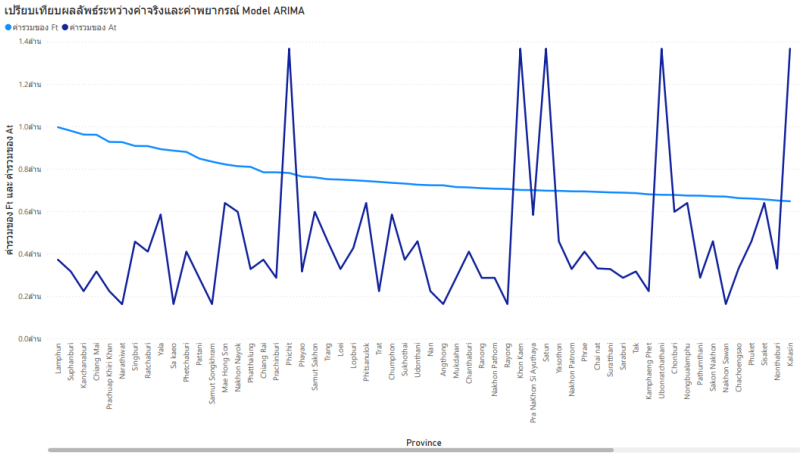
ArimaModel

```
Forecast Model trained on the following time series:
Name of time series: Unemployed Number of values: 100

Resulting Forecast Model:
Arima Model (p: 1,d: 0, q: 1)
AR Coefficients: [0.24994685532511082]
MA Coefficients: [0.14244586434566875]
constant: 5089.100019489319, length of residuals: 1
```

ภาพที่ 1 แสดงผลลัพธ์ของการพยากรณ์โดยใช้แบบจำลอง ARIMA

ผลลัพธ์ของแบบจำลอง ARIMA โดยใช้ตัวแปร (Attribute) ผู้ว่างงาน (Unemployed) ในการพยากรณ์ ค่า AR Coefficients คือการถดถอยอัตโนมัติเท่ากับ 0.2499 หมายความว่าความสัมพันธ์ของข้อมูลในอดีตมีอิทธิพลต่อค่าพยากรณ์อย่างมีนัยสำคัญค่า MA Coefficient คือค่าเฉลี่ยเคลื่อนที่เท่ากับ 0.1424 หมายความว่าข้อผิดพลาดในอดีตส่งผลกระทบอย่างมากต่อค่าพยากรณ์ในปัจจุบัน ค่า constant คือค่าคงที่เท่ากับ 5089.10 หมายความว่า ข้อมูลส่วนใหญ่อยู่ใกล้เคียงกับค่านี้ ซึ่งบ่งบอกถึงความเหมาะสมของการพยากรณ์ในแบบจำลอง ARIMA สามารถอธิบายแนวโน้มของจำนวนผู้ว่างงานได้อย่างมีนัยสำคัญผ่านองค์ประกอบของความสัมพันธ์เชิงเวลาและข้อผิดพลาดในอดีต



ภาพที่ 2 ผลการเปรียบเทียบค่าพยากรณ์และค่าจริงรายจังหวัดด้วยแบบจำลอง ARIMA

จากภาพที่ 2 แสดงการเปรียบเทียบค่าพยากรณ์และค่าจริงของจำนวนผู้ว่างงานในแต่ละจังหวัดของประเทศไทยทั้ง 77 จังหวัด โดยใช้โปรแกรม Power BI สำหรับการแสดงผล และใช้ RapidMiner ในการพยากรณ์ข้อมูลด้วยแบบจำลอง ARIMA กราฟเส้นนี้ประกอบด้วยแกน X ซึ่งแสดงรายชื่อจังหวัดในประเทศไทย แกน Y แสดงค่าจำนวนผู้ว่างงาน โดยเส้นสีฟ้าอ่อนแทนค่าพยากรณ์ และเส้นสีน้ำเงินเข้มแทนค่าจริงจากกราฟพบว่า ค่าพยากรณ์มีลักษณะเรียบและลดลงอย่างต่อเนื่องในหลายจังหวัด ขณะที่ค่าจริงมีความผันผวนสูง และมีบางจังหวัดที่ค่าจริงพุ่งขึ้นอย่างชัดเจนความแตกต่างระหว่างเส้นทั้งสองสะท้อนถึงความคลาดเคลื่อนของ

แบบจำลอง โดยเฉพาะในจังหวัดที่มีการเปลี่ยนแปลงจำนวนผู้ว่างงานสูง ทำให้เห็นได้ว่าแบบจำลอง ARIMA อาจไม่สามารถจับแนวโน้มแบบไม่สม่ำเสมอในพื้นที่ต่าง ๆ ได้อย่างแม่นยำ

4.2 แบบจำลองโฮลต์–วินเทอร์ส (Holt–Winters) เป็นวิธีการพยากรณ์ข้อมูลอนุกรมเวลาที่สามารถจัดการกับแนวโน้มและฤดูกาลของข้อมูลได้อย่างมีประสิทธิภาพ (Hyndman & Athanasopoulos, 2021) ผู้วิจัยนำแบบจำลองโฮลต์–วินเทอร์ส (Holt–Winters) มาใช้ในการวิเคราะห์และพยากรณ์อัตราการว่างงานของประเทศไทยในช่วงปี พ.ศ. 2564–2567 โดยดำเนินการวิเคราะห์ด้วยโปรแกรม RapidMiner Studio เพื่อเปรียบเทียบประสิทธิภาพกับแบบจำลองอื่น

Level (ค่าแนวโน้มระดับพื้นฐาน)

$$Lt = \alpha(Yt - St - s) + (1 - \alpha)(Lt - 1 + Tt - 1)$$

Trend (แนวโน้ม)

$$Tt = \beta(Lt - Lt - 1) + (1 - \beta)Tt - 1$$

Seasonal (ฤดูกาล)

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}$$

Forecast (การพยากรณ์ล่วงหน้า)

$$\hat{Y}_{t+m} = L_t + mT_t + S_{t-ts+m \bmod s}$$

โดยที่

Yt ค่าของตัวแปร ณ เวลา t

$\hat{Y}_{t+m}$ ค่าพยากรณ์ที่ระยะเวลา $t + m$

L_t ค่า level ณ เวลาt

T_t ค่าความเปลี่ยนแปลงแนวโน้ม

S_tค่าฤดูกาล

α, β, γ ค่าคงที่ (0–1) สำหรับการปรับระดับ แนวโน้ม และฤดูกาล

s ความยาวของรอบฤดูกาล (เช่น 12 เดือน)

m ระยะเวลาที่ต้องการพยากรณ์ล่วงหน้า

HoltWintersModel

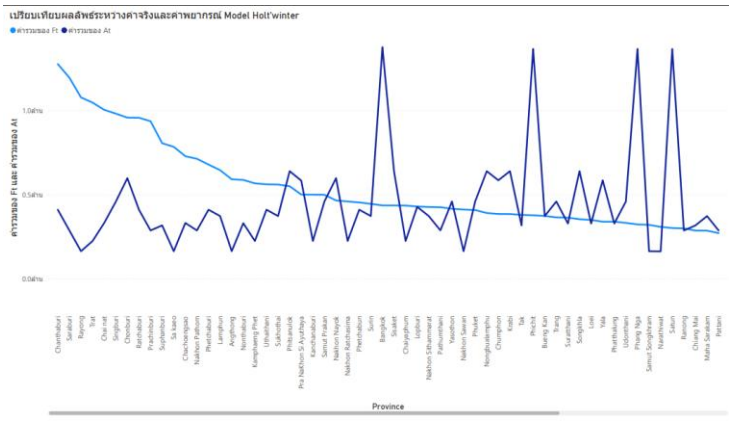
```
Forecast Model trained on the following time series:
Name of time series: Unemployed Number of values: 100

Resulting Forecast Model:
Holt-Winters Model (alpha: 0.5, beta: 0.1, gamma: 0.5)
Period: 3, mode type: MULTIPLICATIVE
```

ภาพที่ 3 แสดงผลลัพธ์ของการพยากรณ์โดยใช้แบบจำลอง Holt’s winter

ผลลัพธ์ของแบบจำลอง Holt’s winter โดยใช้ตัวแปร (Attribute) ผู้รอฤดูกาล (Waiting for work) ในการพยากรณ์ ค่า Alpha คือค่าที่ควบคุมน้ำหนักของข้อมูลในอดีตเท่ากับ 0.5 หมายความว่าแบบจำลองให้ความสำคัญกับข้อมูลปัจจุบันมากกว่าอดีต ค่า beta คือค่าที่ควบคุมแนวโน้มเท่ากับ 0.1 หมายความว่า

แบบจำลองปรับแนวโน้มช้าและค่าgamma คือค่าที่ควบคุมฤดูกาลเท่ากับ 0.5 หมายความว่าแบบจำลองปรับฤดูกาลเร็ว



ภาพที่ 4 ผลการเปรียบเทียบค่าพยากรณ์และค่าจริงรายจังหวัดด้วยแบบจำลอง Holt–Winters

จากภาพที่ 4 เป็นการเปรียบเทียบค่าพยากรณ์และค่าจริงของจำนวนผู้ว่างงานในแต่ละจังหวัดของประเทศไทยทั้ง 77 จังหวัด โดยใช้โปรแกรม Power BI แสดงผล และใช้ Rapid Miner พยากรณ์ข้อมูลด้วยแบบจำลอง Holt's Winters ซึ่งในกราฟ แกน X แสดงรายชื่อจังหวัด แกน Y หลักแสดงค่าพยากรณ์ (เส้นสีน้ำเงินอ่อน) และแกน Y รองแสดงค่าจริง (เส้นสีน้ำเงินเข้ม) ผลลัพธ์แสดงให้เห็นว่าแบบจำลองสามารถพยากรณ์ได้ใกล้เคียงกับค่าจริงในบางจังหวัด แต่ในหลายพื้นที่พบความคลาดเคลื่อนอย่างชัดเจน โดยเฉพาะในจังหวัดที่มีความผันผวนสูง ค่าพยากรณ์มีลักษณะราบเรียบ ต่างจากค่าจริงที่มีความแปรปรวน ซึ่งสะท้อนว่าแบบจำลอง Holt's Winters อาจไม่สามารถจับลักษณะข้อมูลในบางพื้นที่ได้อย่างแม่นยำ

4.3 แบบจำลองเคเนียเรสต์เนเบอร์ส (K-NN) เป็นแบบจำลองการเรียนรู้ของเครื่องที่อาศัยความใกล้เคียงของข้อมูลเพื่อใช้ในการจำแนกหรือพยากรณ์ผลลัพธ์ โดยไม่ต้องกำหนดสมมติฐานเชิงเส้นของข้อมูลล่วงหน้า (Cover & Hart, 1967) จุดเด่นของแบบจำลองนี้คือไม่ต้องมีขั้นตอนการฝึกแบบจำลองล่วงหน้า (Lazy Learning) แต่จะใช้ข้อมูลในอดีตในการคำนวณระยะห่างเพื่อค้นหาข้อมูลที่มีความใกล้เคียงมากที่สุดในการตัดสินผลลัพธ์ ผู้วิจัยนำแบบจำลองเคเนียเรสต์เนเบอร์ส (K-NN) มาใช้ในการวิเคราะห์และพยากรณ์อัตราการว่างงานของประเทศไทยในช่วงปี พ.ศ. 2564–2567 โดยดำเนินการวิเคราะห์ด้วยโปรแกรม RapidMiner Studio ในการวิเคราะห์ข้อมูล

ในการทำนายค่าเป้าหมาย $\hat{y}$ สำหรับจุดข้อมูลใหม่

1. คำนวณระยะห่างระหว่าง x กับจุดข้อมูลอื่น ๆ ในชุดข้อมูลฝึก

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2}$$

2. เลือก K จุดข้อมูลที่มีระยะใกล้ที่สุดกับ x
3. นำค่าเป้าหมายของ K จุดนั้นมาหาค่าเฉลี่ย (ในกรณี regression)

$$\hat{y} = \frac{1}{K} \sum_{i \epsilon Nk(x)}^{n} y_i$$

โดยที่

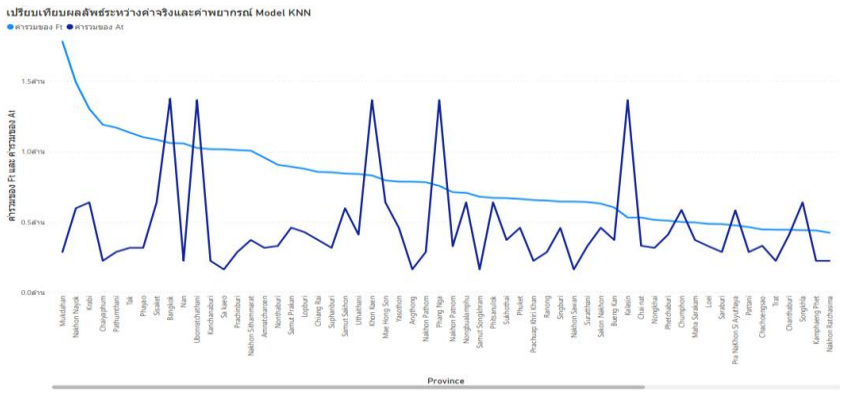
$Nk(x)$ กลุ่มของ K ตัวอย่างที่ใกล้ที่สุดกับ x

```
KNNRegression

Weighted 5-Nearest Neighbour model for regression.
The model contains 3696 examples with 7 dimensions.
```

ภาพที่ 5 แสดงผลลัพธ์ ของการพยากรณ์โดยใช้แบบจำลอง K-NN

ผลลัพธ์ของแบบจำลอง K-NN โดยใช้ตัวแปร (Attribute) ผู้ว่างงาน (Unemployee) ในการพยากรณ์ เมื่อทำการ Cross validation แล้วได้ลัพธ์เท่ากับ 7 dimensions ซึ่งเป็นค่าใกล้เคียง 7 จังหวัดที่มีค่าพยากรณ์และค่าจริงที่มีความใกล้เคียงกัน ได้แก่ จังหวัด แม่ฮ่องสอน หนองบัวลำภู พิษณุโลก ชุมพร จันทบุรี เพชรบุรี และ พระนครศรีอยุธยา



ภาพที่ 6 ผลการเปรียบเทียบค่าพยากรณ์และค่าจริงรายจังหวัดด้วยแบบจำลอง K-NN

จากภาพที่ 6 แสดงการเปรียบเทียบค่าพยากรณ์และค่าจริงของจำนวนผู้ว่างงานในแต่ละจังหวัดของประเทศไทยทั้ง 77 จังหวัด โดยใช้โปรแกรม Power BI สำหรับการแสดงผลและใช้ RapidMiner ในการพยากรณ์ข้อมูลด้วยแบบจำลอง K-NN กราฟแสดงรายชื่อจังหวัดบนแกน X และจำนวนผู้ว่างงานบนแกน Y โดยเส้นสีฟ้าอ่อนคือค่าพยากรณ์ (Ft) และเส้นสีน้ำเงินเข้มคือค่าจริง (At) จากกราฟจะเห็นว่า ค่าพยากรณ์ของ KNN มีแนวโน้มลดลงเช่นเดียวกับ ARIMA แต่มีลักษณะลาดเอียงมากกว่า และในบางจังหวัดแบบจำลองสามารถพยากรณ์ใกล้เคียงค่าจริงได้อย่างไรก็ตาม ในหลายพื้นที่ยังพบความคลาดเคลื่อนอย่างชัดเจน โดยเฉพาะในจังหวัดที่มีค่าจริงพุ่งสูงแบบจำลองยังไม่สามารถตอบสนองต่อความเปลี่ยนแปลงแบบเฉียบพลัน

ได้ดีนักซึ่งสะท้อนว่า KNN อาจมีประสิทธิภาพที่ดีกว่า ARIMA เล็กน้อยในบางกรณี แต่ยังไม่สามารถจัดการกับข้อมูลที่ผันผวนสูงได้อย่างแม่นยำทั่วทั้งประเทศ

4.4 การถดถอยเชิงเส้น (Linear Regression) เป็นแบบจำลองทางสถิติพื้นฐานที่ใช้ในการอธิบายและวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยอาศัยสมมติฐานความสัมพันธ์เชิงเส้นของข้อมูล และมักถูกนำมาใช้เป็นแบบจำลองอ้างอิงพื้นฐาน (Baseline Model) ในการเปรียบเทียบประสิทธิภาพกับแบบจำลองการพยากรณ์รูปแบบอื่น (Montgomery et al., 2012) สามารถนำมาใช้เพื่อพยากรณ์ค่าหรืออธิบายความสัมพันธ์ ดังนี้

สมการ Linear Regression (Simple Linear Regression)

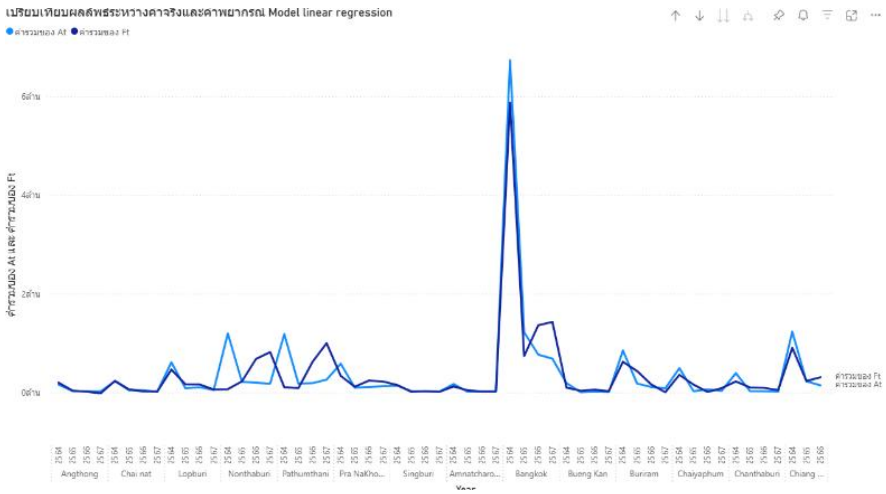
$$\hat{y} = \beta_0 + \beta_1 x$$

$\hat{y}$ ค่าที่พยากรณ์หรือประมาณจากโมเดล

x ตัวแปรต้น (Independent variable)

β_0 ค่าคงที่ (Intercept)

β_1 ค่าสัมประสิทธิ์ของ x (Slope)



ภาพที่ 7 ผลการเปรียบเทียบค่าพยากรณ์และค่าจริงรายจังหวัดด้วย Linear Regression

แสดงการเปรียบเทียบค่าพยากรณ์และค่าจริงของจำนวนผู้ว่างงานในประเทศไทย โดยพิจารณาข้อมูลเชิงเวลาในช่วงปี 2564–2567 แบ่งตามจังหวัด โดยใช้โปรแกรม Power BI ในการแสดงผล และใช้ RapidMiner สร้างแบบจำลองพยากรณ์ด้วย Linear Regressionแกน X แสดงข้อมูลจังหวัดและปีควบคู่กัน เช่น "Bangkok 2564", "Lopburi 2565" เป็นต้น ส่วนแกน Y แสดงจำนวนผู้ว่างงาน โดยเส้นสีน้ำเงินเข้มคือค่าพยากรณ์ (Ft) และเส้นฟ้าอ่อนคือค่าจริง (At) จากกราฟจะเห็นว่าแบบจำลอง Linear Regression สามารถจับแนวโน้มของข้อมูลได้ดีกว่าแบบจำลองอื่น โดยมีหลายจุดที่ค่าพยากรณ์มีความใกล้เคียงกับค่าจริง โดยเฉพาะในปีหลัง ๆ ถึงแม้จะมีบางช่วงที่ค่าพยากรณ์ยังแตกต่างจากค่าจริงอยู่บ้าง แต่ภาพรวมแสดงว่า Linear Regression มีแนวโน้มพยากรณ์ได้แม่นยำและสอดคล้องกับข้อมูลจริงมากกว่า ARIMA และ KNN จึงอาจสรุปได้ว่า Linear Regression เป็นแบบจำลองที่มีประสิทธิภาพดีกว่าในการพยากรณ์จำนวนผู้ว่างงานในเชิงพื้นที่และเวลา

5. การวัดประสิทธิภาพของแบบจำลอง (Evaluation) การทดสอบ ประสิทธิภาพของแบบจำลองได้ดำเนินการวัดค่า RMSE ค่านี้จะให้ความสำคัญกับค่าผิดพลาดที่มาก โดยค่า RMSE ที่ต่ำแสดงว่าแบบจำลองมีความแม่นยำสูง ,MAE ค่านี้จะให้ความสำคัญกับขนาดของความคลาดเคลื่อนโดยรวม ค่า MAE ที่ต่ำแสดงว่าแบบจำลองมีความแม่นยำสูง MAPE วัดความแม่นยำของแบบจำลองพยากรณ์

สูตรการคำนวณ MAPE

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

โดยที่

A_t = ค่าจริง (Actual value)

F_t = ค่าที่พยากรณ์ (Forecast value)

n = จำนวนข้อมูลทั้งหด

ตารางที่ 1 การเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง

Model	MAPE	RMSE	MAE	ความหมาย
ARIMA	11%	8578.801	6658.515	โมเดล ARIMA มีความแม่นยำปานกลาง ดีสำหรับจับแนวโน้มโดยรวม แต่บางค่าคลาดเคลื่อนมาก
Holt's winters	4%	11534.302	9045.802	โมเดล Holt's Winters แม่นยำเชิงสัมพัทธ์สูง ดีสำหรับข้อมูลที่มีแนวโน้มและฤดูกาล แต่ตัวเลขจริงยังคลาดเคลื่อนบ้าง
K-NN	6%	62036.073	8556.119	โมเดล K-Nearest Neighbors (K-NN) มีความแม่นยำเชิงสัมพัทธ์ดี (MAPE ต่ำ) แต่ RMSE สูงมาก แปลว่ามีบางค่าผิดพลาดสูง เหมาะกับแนวโน้มทั่วไป
Linear regression	0%	344393.751	15617.772	โมเดลนี้มีค่า MAPE ต่ำสุด แสดงว่าจับแนวโน้มเชิงสัมพัทธ์ได้ดีมาก แต่ RMSE สูงสุด แสดงว่าตัวเลขจริงบางค่าผิดพลาดมากมีค่า MAPE ต่ำมาก อาจเป็นสัญญาณของการเรียนรู้เกินชุดข้อมูล (overfitting)

จากผลการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 4 แบบ พบว่าแบบจำลอง ARIMA มีค่า MAPE เท่ากับ ร้อยละ 11 จัดอยู่ในระดับความแม่นยำปานกลาง แต่มีจุดเด่นคือค่า RMSE (8,578) และ MAE (6,659) ต่ำที่สุดเมื่อเทียบกับทุกแบบ แสดงให้เห็นว่าความคลาดเคลื่อนเชิงตัวเลขอยู่ในระดับต่ำและมีความเสถียรสูง ขณะที่แบบจำลอง Holt's Winters ให้ค่า MAPE ต่ำที่สุดเพียงร้อยละ 4 ซึ่งบ่งบอกถึงความแม่นยำเชิงสัมพัทธ์ที่ดีที่สุด แต่ค่าความคลาดเคลื่อนเชิงตัวเลขกลับสูงกว่า ARIMA (RMSE = 11,534 และ MAE

= 9,046) จึงอาจเหมาะกับข้อมูลที่มีแนวโน้มและฤดูกาลชัดเจนมากกว่าข้อมูลที่มีความผันผวน ส่วนแบบจำลอง K-Nearest Neighbors (K-NN) แม้จะมีค่า MAPE เท่ากับร้อยละ 6 ซึ่งดีกว่า ARIMA แต่กลับมีค่า RMSE สูงถึง 62,036 และ MAE เท่ากับ 8,556 บ่งบอกถึงความไม่เสถียรและการกระจายตัวของค่าความคลาดเคลื่อนในระดับสูง ขณะที่การถดถอยเชิงเส้น (Linear Regression) แม้ค่า MAPE จะเท่ากับร้อยละ 0 ซึ่งดูเหมือนมีความแม่นยำสูงที่สุด แต่ค่าความคลาดเคลื่อนเชิงตัวเลขกลับสูงมาก (RMSE = 344,394 และ MAE = 15,618) สะท้อนถึงปัญหาการเรียนรู้เกินชุดข้อมูล (Overfitting) และความไม่เหมาะสมของแบบจำลองเชิงเส้นต่อข้อมูลชุดนี้ ดังนั้น เมื่อพิจารณาตัวชี้วัดทั้งหมดร่วมกัน แบบจำลอง ARIMA จึงเป็นแบบจำลองที่เหมาะสมที่สุด เนื่องจากสามารถสร้างสมดุล

ผลการวิจัย

ผลการประเมินประสิทธิภาพของแบบจำลองแต่ละแบบพบว่าแบบจำลอง ARIMA ให้ค่าความคลาดเคลื่อนเชิงตัวเลขต่ำที่สุด โดยมีค่า RMSE และ MAE ต่ำกว่าแบบจำลองอื่น สะท้อนถึงความเสถียรของการพยากรณ์ในเชิงปริมาณ แบบจำลอง Holt's Winters ให้ค่า MAPE ต่ำที่สุด แสดงถึงความแม่นยำเชิงเปอร์เซ็นต์ที่ดี แต่ยังมีค่าความคลาดเคลื่อนเชิงตัวเลขสูงกว่า ARIMA ขณะที่แบบจำลอง K-Nearest Neighbors (K-NN) แม้จะให้ค่า MAPE อยู่ในระดับดี แต่มีค่า RMSE สูงมาก บ่งบอกถึงความไม่เสถียรของผลการพยากรณ์ ส่วนแบบจำลอง Linear Regression ให้ค่า MAPE ต่ำผิดปกติ แต่มีค่า RMSE และ MAE สูงที่สุด สะท้อนถึงปัญหาความไม่เหมาะสมของแบบจำลองกับข้อมูลชุดนี้

จากการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ทั้ง 4 แบบ ได้แก่ ARIMA, Holt's Winters, K-Nearest Neighbors (K-NN) และ Linear Regression โดยใช้ตัวชี้วัดความแม่นยำ 3 ค่า คือ ค่าร้อยละความคลาดเคลื่อนค่าเฉลี่ย(MAPE), ค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (RMSE) และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAE) พบว่าแบบจำลองแต่ละแบบมีจุดเด่นและข้อจำกัดแตกต่างกัน แบบจำลอง ARIMA มีค่า MAPE เท่ากับ 11% จัดอยู่ในระดับความแม่นยำปานกลาง แต่มีค่า RMSE และ MAE เท่ากับ 8,578.801 และ 6,658.515 ตามลำดับ ซึ่งเป็นค่าความคลาดเคลื่อนเชิงตัวเลขที่ต่ำที่สุดในบรรดาทุกโมเดล แสดงให้เห็นว่าแบบจำลอง ARIMA สามารถทำนายได้ใกล้เคียงค่าจริงโดยเฉลี่ยมากที่สุด แบบจำลอง Holt's Winters ให้ค่า MAPE เท่ากับ 4% ซึ่งเป็นค่าที่ต่ำที่สุด หมายความว่ามีความแม่นยำเชิงเปอร์เซ็นต์สูงที่สุด แต่ค่าความคลาดเคลื่อนเชิงตัวเลข (RMSE = 11,534.302 และ MAE = 9,045.802) ยังสูงกว่า ARIMA ทำให้แม้จะมีความแม่นยำเชิงสัมพัทธ์ที่ดี แต่ค่าผิดพลาดในเชิงปริมาณยังคงมากกว่า แบบจำลอง K-NN มีค่า MAPE เท่ากับ 6% ซึ่งดีกว่า ARIMA แต่มีค่า RMSE และ MAE เท่ากับ 62,036.073 และ 8,556.119 ตามลำดับ โดยเฉพาะค่า RMSE ที่สูงมาก บ่งบอกถึงความไม่เสถียรและการกระจายของค่าความคลาดเคลื่อนที่กว้าง จึงไม่เหมาะสมสำหรับการนำไปใช้งานจริง และการถดถอยเชิงเส้น (Linear Regression) ให้ค่า MAPE เท่ากับ 0% ซึ่งดูเหมือนจะแม่นยำที่สุด แต่ค่าความคลาดเคลื่อนเชิงตัวเลข (RMSE = 344,393.751 และ MAE = 15,617.772) กลับสูงที่สุดในบรรดาทุกแบบจำลอง โดยสรุปผลการวิจัยชี้ให้เห็นว่าแบบจำลอง ARIMA เป็นแบบจำลองที่เหมาะสมที่สุด เนื่องจากมีค่า RMSE และ MAE ต่ำที่สุด ทำให้สามารถพยากรณ์ได้ใกล้เคียงค่าจริงมากกว่าแบบจำลองอื่น ในขณะที่ Holt's Winters แม้จะมีความแม่นยำเชิงเปอร์เซ็นต์สูงกว่า แต่มีค่าความคลาดเคลื่อนเชิงตัวเลขมากกว่าเล็กน้อย จึงเป็นตัวเลือกที่น่าสนใจเป็น

อันดับรองลงมา ส่วน K-NN และ Linear Regression มีข้อจำกัดด้านความแม่นยำและค่าความคลาดเคลื่อน จึงไม่เหมาะสมสำหรับการนำมาใช้ในงานวิจัยนี้

อภิปรายผล

ผลการวิจัยครั้งนี้ชี้ให้เห็นว่า ARIMA เป็นแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์ เนื่องจากมีค่า RMSE (8,578.801) และ MAE (6,658.515) ต่ำที่สุดในทุกแบบจำลอง สะท้อนว่าผลการพยากรณ์มีความใกล้เคียงกับค่าจริงมากที่สุด แม้ว่าค่า MAPE จะอยู่ที่ 11% ซึ่งไม่ใช่ค่าที่ต่ำที่สุด แต่เมื่อพิจารณาความสมดุลของตัวชี้วัดทั้งหมดแล้ว ARIMA แสดงให้เห็นถึงประสิทธิภาพที่ดีกว่าแบบจำลองอื่นซึ่งสอดคล้องกับสำนักงานสถิติแห่งชาติ (2566) ที่ระบุว่าแบบจำลอง ARIMA มักให้ผลการพยากรณ์ที่มีความแม่นยำสูงในข้อมูลอนุกรมเวลาที่มีแนวโน้มชัดเจน และสอดคล้องกับงานวิจัยของวรางคณา กีรติวิบูลย์ (2559) ที่พบว่าการใช้ ARIMA เป็นหนึ่งในองค์ประกอบของการพยากรณ์รวมช่วยเพิ่มประสิทธิภาพในการพยากรณ์ข้อมูลผู้ว่างงานของประเทศไทย เมื่อเทียบกับ Holt's Winters แม้จะให้ค่า MAPE ต่ำที่สุดเพียง 4% ซึ่งแสดงถึงความแม่นยำเชิงเปอร์เซ็นต์สูง แต่กลับมีค่า RMSE (11,534.302) และ MAE (9,045.802) สูงกว่า ARIMA ทำให้มีความคลาดเคลื่อนเชิงปริมาณมากกว่า จึงอาจไม่เหมาะสมหากต้องการผลลัพธ์ที่ใกล้เคียงกับค่าจริงในเชิงตัวเลข ความต่างนี้สอดคล้องกับข้อสังเกตในงานวิจัยของวรางคณา กีรติวิบูลย์ (2559) ที่พบว่า ARIMA กับ Holt–Winters สามารถลดข้อด้อยของวิธีเดี่ยวได้ชี้ทางต่อยอดว่าการถ่วงน้ำหนักการพยากรณ์อาจยกระดับผลลัพธ์ได้ในบริบทข้อมูล สำหรับ K-Nearest Neighbors (K-NN) พบว่ามีค่า MAPE เท่ากับ 6% ซึ่งดีกว่า ARIMA แต่ค่าความคลาดเคลื่อนเชิงปริมาณกลับสูงมากโดยเฉพาะค่า RMSE สูงถึง 62,036.073 บ่งบอกว่าผลการพยากรณ์มีความผันผวนสูงและไม่เสถียร จึงไม่สามารถใช้เป็นแบบจำลองหลักได้ในขณะที่ Linear Regression แม้จะให้ค่า MAPE = 0% ซึ่งดูเหมือนว่าจะทำนายได้แม่นยำที่สุด แต่กลับมีค่า RMSE และ MAE สูงถึง 344,393.751 และ 15,617.772 ตามลำดับ แสดงให้เห็นถึงปัญหาการเรียนรู้เกินชุดข้อมูล (Overfitting) หรือความไม่เหมาะสมของแบบจำลองเชิงเส้นในการอธิบายข้อมูลอนุกรมเวลา ทำให้ผลการพยากรณ์บิดเบือนไปจากค่าจริงอย่างมาก

ผลการวิจัยพบว่าแบบจำลองที่เหมาะสมที่สุดคือ ARIMA เนื่องจากสามารถสร้างสมดุลระหว่างความแม่นยำเชิงสัมพัทธ์และเชิงปริมาณได้ดีที่สุด ซึ่งสอดคล้องกับงานของสำนักงานสถิติแห่งชาติ (2566) ที่ระบุว่า ARIMA มักให้ผลการพยากรณ์ที่มีความแม่นยำสูงในข้อมูลอนุกรมเวลาที่มีแนวโน้มชัดเจน และผลการศึกษานี้ยังสอดคล้องกับงานวิจัยของ Mahipan et al. (2013) ที่พบว่าแบบจำลองอารีมา (ARIMA) ให้ผลการพยากรณ์อัตราการว่างงานของประเทศไทยที่มีความแม่นยำสูงเมื่อเปรียบเทียบกับแบบจำลองอื่น ผลการวิจัยชี้ว่า อัตราการว่างงานของไทยมีความผันผวนในบางเดือน โดยเฉพาะหลังโควิด-19 และช่วงที่การท่องเที่ยวฟื้นตัว แบบจำลองที่สามารถอธิบายฤดูกาลและแนวโน้มของข้อมูลได้จะให้ผลการพยากรณ์ที่มีประสิทธิภาพมากกว่า ซึ่งสอดคล้องกับปัจจัยกำหนดอัตราการว่างงานที่นำเสนอในงานวิจัยของ Khanthavit (2021) และ Zhou (2023) ซึ่งชี้ให้เห็นถึงบทบาทของโครงสร้างเศรษฐกิจและปัจจัยเชิงมหภาคที่ส่งผลต่อความผันผวนของตลาดแรงงาน เช่น การเตรียมมาตรการรองรับแรงงานในช่วงนอกฤดูกาล การออกนโยบายส่งเสริมการจ้างงานระยะสั้น หรือการวางแผนพัฒนากำลังคนให้สอดคล้องกับการเติบโตของภาคท่องเที่ยวและบริการ

สรุปผลการวิจัยนี้มุ่งศึกษาประสิทธิภาพและเปรียบเทียบความเหมาะสมของแบบจำลองการพยากรณ์อัตราการว่างงานในประเทศไทย ได้แก่ ARIMA, Holt's Winters, K-Nearest Neighbors (K-NN) และ

Linear Regression โดยใช้ตัวชี้วัดความแม่นยำ ได้แก่ MAPE, RMSE และ MAE ผลการวิจัยพบว่าแบบจำลอง ARIMA มีความเหมาะสมที่สุด เนื่องจากให้ค่าความคลาดเคลื่อนเชิงตัวเลขต่ำที่สุด ขณะที่ Holt's Winters มีความแม่นยำเชิงเปอร์เซ็นต์สูง แต่มีค่าความคลาดเคลื่อนเชิงตัวเลขสูงกว่า ส่วนแบบจำลอง K-NN และ Linear Regression ยังมีข้อจำกัดด้านความเสถียรและความเหมาะสมกับข้อมูลอนุกรมเวลา

ข้อเสนอแนะ

จากการศึกษาครั้งนี้พบว่าข้อมูลที่ใช้ในการวิเคราะห์มีข้อจำกัดบางประการ โดยเฉพาะช่วงเวลาและลักษณะข้อมูลที่มีความผันผวนสูง ซึ่งอาจทำให้แบบจำลองบางชนิดไม่สามารถสะท้อนการเปลี่ยนแปลงได้อย่างแม่นยำ ควรใช้ข้อมูลหลายช่วง / หลายแหล่ง และรวมปัจจัยภายนอก (ตัวแปรเศรษฐกิจ/ฤดูกาล) เพื่อเพิ่มความครอบคลุมของข้อมูลที่มีผลต่อการว่างงาน

เอกสารอ้างอิง

กองบริหารตลาดแรงงาน. (2567). *สถิติความต้องการแรงงานรายจังหวัด*. สืบค้น 28 สิงหาคม 2567. https://www.doe.go.th

ฐาปณีย์ บุญชอบ. (2563). *เข้าใจ CRISP-DM ฉบับเร่งรัด*. สืบค้น 20 มีนาคม 2567. https://kamboonchob.medium.com

วรางคณา กีรติวิบูลย์. (2559). การพยากรณ์จำนวนผู้ว่างงานในประเทศไทย. *Asian Health, Science and Technology Reports (AHSTR)*. 24(1), 102–114. https://ph03.tci-thaijo.org/index.php/ahstr/article/view/1893

สำนักงานสถิติแห่งชาติ. (2566). *รายงานการศึกษาค่าคาดการอัตราการว่างงานของประเทศไทย*. สืบค้น 28 สิงหาคม 2567. https://www.nso.go.th/public/e-book/Analytical-Reports/Report_Unemployed_2566/66/

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 13(1), 21–27. https://ieeexplore.ieee.org/document/1053964

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

Khanthavit, A. (2021). A causality analysis of lottery ambling and Unemployment in Thailand. *Journal of Asian Finance Economics and Business*. 8(8), 149–156. https://www.koreascience.kr/article/JAKO202120953711352.pdf

Mahipan, K., Chutiman, N., & Kumphon, B. (2013). A forecasting model for Thailand's unemployment rate. *Modern Applied Science*. 7(7), 10–16. https://doi.org/10.5539/mas.v7n7p10

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (5th ed.). Wiley.

Zhou, Z. (2023). Research of the influencing factors on unemployment rate. *Highlights in Business, Economics and Management*. 5, 134–141. https://doi.org/10.54097/hbem.v5i.5040