การประยุกต์ใช้ Google Trends ในการประมาณการอัตราการว่างงานในประเทศไทย

โดย

จิราภา เฮนะเกษตร

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
เศรษฐศาสตรมหาบัณฑิต
สาขาเศรษฐศาสตร์ธุรกิจ
คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2564

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

Ref. code: 25646304010116LCN

ESTIMATING THAILAND'S UNEMPLOYMENT RATE: AN APPLICATION OF GOOGLE TRENDS DATA

BY

JIRAPA HANAKASET

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS
BUSINESS ECONOMICS
FACULTY OF ECONOMICS
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2021

COPYRIGHT OF THAMMASAT UNIVERSITY

Ref. code: 25646304010116LCN

มหาวิทยาลัยธรรมศาสตร์
คณะเศรษฐศาสตร์

การค้นคว้าอิสระ

ของ

จิราภา เฮนะเกษตร

เรื่อง

การประยุกต์ใช้ Google Trends ในการประมาณการอัตราการว่างงานในประเทศไทย

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
เศรษฐศาสตรมหาบัณฑิต (เศรษฐศาสตร์ธุรกิจ)

เมื่อ วันที่ 25 กรกฎาคม พ.ศ. 2565

อาจารย์ที่ปรึกษาการค้นคว้าอิสระ

(รองศาสตราจารย์ ดร.ณัฐพงษ์ พัฒนพงษ์)

กรรมการสอบการค้นคว้าอิสระ

(ผู้ช่วยศาสตราจารย์ ดร.วศิน ศิวสฤษดิ์)

คณบดี

(ผู้ช่วยศาสตราจารย์ ดร.ศุภชัย ศรีสุชาติ)

หัวข้อการค้นคว้าอิสระ	การประยุกต์ใช้ Google Trends ในการประมาณการอัตราการว่างงานในประเทศไทย
ชื่อผู้เขียน	จิราภา เฮนะเกษตร
ชื่อปริญญา	เศรษฐศาสตรมหาบัณฑิต
สาขาวิชา/คณะ/มหาวิทยาลัย	เศรษฐศาสตร์ธุรกิจ เศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	รองศาสตราจารย์ ดร.ณัฐพงษ์ พัฒนพงษ์
ปีการศึกษา	2564

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อใช้ข้อมูลคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends สำหรับการพยากรณ์อัตราการว่างงานในประเทศไทย ด้วยวิธีการเรียนรู้ของ Machine Learning ได้แก่ Linear Regression Model, Optimizable Tree Regression Model, Optimizable Support Vector Regression Model, Neural Network Model และ Optimizable Ensemble Model โดยในการศึกษานี้ผู้ศึกษาใช้ข้อมูลอนุกรมเวลาแบบรายเดือนที่แบ่งออกเป็น 2 แบบจำลอง กล่าวคือ ตัวแปรตามประกอบด้วย 2 ชุดข้อมูล คือ ข้อมูลอัตราการว่างงานที่มาจากสำนักงานสถิติแห่งชาติ ตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนธันวาคม พ.ศ. 2563 และข้อมูลอัตราการว่างงานที่มาจากผู้ประกันตนในประกันสังคม ตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนตุลาคม พ.ศ. 2564 ขณะที่ใช้ตัวแปรอิสระคือคำค้นหาใน Google Trends และแบ่งชุดข้อมูลออกเป็น 80:20 สำหรับข้อมูลที่ใช้ในการเรียนรู้เพื่อสร้างแบบจำลอง และข้อมูลที่ใช้ทดสอบความประสิทธิภาพของแบบจำลอง ผลการศึกษา พบว่า เมื่อพิจารณาจากค่าความคาดเคลื่อนรากที่สองกำลังสองเฉลี่ยต่ำที่สุด วิธีการ Ensemble มีประสิทธิภาพสูงสุดในการพยากรณ์อัตราการว่างงานของสำนักงานสถิติแห่งชาติ ในขณะที่วิธีการ Linear Regression มีประสิทธิภาพสูงสุดในการพยากรณ์อัตราการว่างงานของสำนักงานประกันสังคม ซึ่งการศึกษานี้ สามารถพัฒนาต่อยอดในอนาคตสำหรับการประยุกต์ใช้ Machine Learning กับข้อมูล Google Trends เพื่อการพยากรณ์อัตราการว่างงานในประเทศไทย

คำสำคัญ: การพยากรณ์, Google Trends, อัตราการว่างงาน, Machine Learning

Ref. code: 25646304010116LCN

Independent Study Title	ESTIMATING THAILAND'S UNEMPLOYMENT RATE: AN APPLICATION OF GOOGLE TRENDS DATA
Author	Jirapa Hanakaset
Degree	Master of Arts
Major Field/Faculty/University	Business Economics Economics Thammasat University
Independent Study Advisor	Associate Professor Nattapong Puttanapong, Ph.D
Academic Year	2021

ABSTRACT

This study obtained the statistics of keyword search from Google Trends for forecasting unemployment in Thailand. A set of machine learning methods was employed, comprised of linear regression, optimizable tree regression, optimizable support vector regression, neural network and optimizable ensemble. The official monthly unemployment data from National Statistical Office (NSO) and Social Security Office (SSO) were used as the dependent variables. The NSO statistics cover the period January 2011 - December 2020, while the coverage of SSO data is from January 2011 to October 2021. Technically, the data were divided into 80:20 for training and testing sets. Using the root mean square error (RMSE) as the criterion, it has been found that the ensemble method was most effective for forecasting NSO unemployment rate dataset modeling, while linear regression method was most accurate method for predicting SSO unemployment statistics. These findings suggest the further improvement of applying machine learning technique with Google Trends data to forecast unemployment in Thailand.

Keywords: Forecast, Google Trends, Unemployment rate, Machine learning.

Ref. code: 25646304010116LCN

กิตติกรรมประกาศ

การค้นคว้าอิสระฉบับนี้สำเร็จลุล่วงได้ด้วยดีจากความอนุเคราะห์และความช่วยเหลือจากหลายฝ่าย โดยเฉพาะอย่างยิ่งจากรองศาสตรจารย์ ดร.ณัฐพงษ์ พัฒนพงษ์ อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ที่เสียสละเวลาในการช่วยเหลือ ให้ความรู้ คำแนะนำ และแนวคิดในการค้นคว้าอิสระนี้ตลอดจนให้ข้อเสนอแนะเพื่อนำมาปรับปรุงแก้ไขต่าง ๆ มาโดยตลอด จนการค้นคว้าอิสระนี้เสร็จสมบูรณ์ และผู้ช่วยศาสตราจารย์ ดร.วศิน ศิวสฤษดิ์ กรรมการการค้นคว้าอิสระ ที่เสียสละเวลาแนะนำและข้อเสนอแนะในการแก้ไขข้อบกพร่องต่าง ๆ ให้การค้นคว้าอิสระฉบับนี้สมบูรณ์ยิ่งขึ้น ผู้ศึกษาขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอขอบคุณคณาจารย์ทุกท่านที่ได้ให้ความรู้จนกระทั่งผู้ศึกษาสามารถใช้ความรู้ที่ได้รับในการทำการค้นคว้าอิสระนี้ และขอขอบคุณเจ้าหน้าที่โครงการที่ได้จัดการเอกสาร และดำเนินการต่าง ๆ ทำให้การค้นคว้าอิสระสำเร็จลุล่วงได้ทันกำหนดการ

สุดท้ายนี้ขอขอบคุณบิดา มารดา ทุกคนในครอบครัว และเพื่อน ๆ ที่คอยสนับสนุน ให้ความช่วยหลือ และให้กำลังใจมาโดยตลอดในระหว่างการค้นคว้าอิสระนี้

ผู้ศึกษาคาดหวังเป็นอย่างยิ่งว่างานค้นคว้าอิสระนี้จะเป็นประโยชน์ต่อผู้ที่สนใจเรื่องเกี่ยวกับการพยากรณ์ ทั้งนี้ หากมีข้อบกพร่องประการใด ผู้ศึกษาขอน้อมรับไว้ และขออภัยมา ณ ที่นี้

จิราภา เฮนะเกษตร

Ref. code: 25646304010116LCN

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(2)
กิตติกรรมประกาศ	(3)
สารบัญตาราง	(9)
สารบัญภาพ	(11)
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการศึกษา	5
1.3 ประโยชน์ที่คาดว่าจะได้รับ	6
1.4 ขั้นตอนการดำเนินงานวิจัย	6
1.5 ขอบเขตการศึกษา	6
1.5.1 ข้อมูลของคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends	7
1.5.2 ข้อมูลอัตราการว่างงาน	8
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	9
2.1 แนวคิดและทฤษฎีที่เกี่ยวข้องกับการมีงานทำและการว่างงาน	9
2.1.1 นิยามของตัวแปรที่ใช้ในการศึกษา	9
2.1.1.1 อัตราการว่างงานจากการสำรวจประชากร	9
2.1.1.2 อัตราการว่างงานจากผู้ประกันตนในประกันสังคมที่ขอรับประโยชน์ทดแทนกรณีว่างงาน	12

Ref. code: 25646304010116LCN

2.1.1.3 Google Trends 13
2.1.2 ทฤษฎีเกี่ยวกับการมีงานทำและการว่างงาน 13
2.1.2.1 ทฤษฎีการว่างงานของสำนักคลาสสิค (The classical Theory of Employment) 13
2.1.2.2 ทฤษฎีการว่างงานของสำนักเคนส์ (John Maynard Keynes) 13
2.1.2.3 ทฤษฎีการว่างงานของสำนักนีโอคลาสิค (Neoclassical school) 14
2.1.2.4 ทฤษฎีแรงจูงใจ (motivation theory) 15
2.1.3 ทฤษฎีเกี่ยวกับพฤติกรรมสารสนเทศ 16
2.1.4. ทฤษฎีการพยากรณ์และการสร้างแบบจำลองด้วย Machine Learning 16
2.1.4.1 แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model) 18
2.1.4.2 แบบจำลองต้นไม้ตัดสินใจ (Optimizable Decision Tree Regression Model) 20
2.1.4.3 แบบจำลองซับพอรตเวกเตอร์รีเกรสชัน (Optimizable Support Vector Regression Model: SVR) 21
2.1.4.4 แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN) 25
2.1.4.5 แบบจำลองแบบเรียงซ้อน (Optimizable Ensemble Model: EM) 26
2.2 วรรณกรรมปริทัศน์ 27
2.2.1 งานวิจัยที่เกี่ยวข้องที่มีการนำ Google Trends มาประยุกต์ใช้กับตัวชี้วัดทางเศรษฐกิจ 28
2.2.2 งานวิจัยที่เกี่ยวข้องที่มีการนำ Google Trends มาประยุกต์ใช้กับตัวแปรอื่น ๆ 34
2.2.3 งานวิจัยที่เกี่ยวข้องที่มีการนำแบบจำลองด้านปัญญาประดิษฐ์มาประยุกต์ใช้ประมาณการตัวแปรต่าง ๆ 37

Ref. code: 25646304010116LCN

บทที่ 3 วิธีการวิจัย 39

3.1 ข้อมูลที่ใช้ในการศึกษา 39
3.1.1 แหล่งข้อมูล 39
3.1.1.1 ข้อมูลตัวแปรต้น หรือแปรอิสระ (Independent Variable) 39
3.1.1.2 ข้อมูลตัวแปรตาม (Dependent Variable) 39
3.1.2 รายละเอียดของข้อมูลที่ใช้ในการศึกษา 40
3.1.2.1 ข้อมูล Google Trends 40
3.1.2.2 ข้อมูลอัตราการว่างงานจากสำนักงานสถิติแห่งชาติ (NSO) 41
3.1.2.3 ข้อมูลอัตราการว่างงานจากผู้ประกันตนในประกันสังคม (SSO) 42
3.1.3 การแบ่งชุดข้อมูลเพื่อทดสอบประสิทธิภาพของแบบจำลอง 43
3.2 กรอบแนวคิดในการศึกษา 45
3.3 แบบจำลองที่ใช้ในการศึกษ 45
3.3.1 แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model) 46
3.3.2 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree Regression Model) 48
3.3.3 แบบจำลองซับพอรตเวกเตอร์รีเกรสชัน (Support Vector Regression Model: SVR) 49
3.3.4 แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN) 51
3.3.5 แบบจำลองแบบเรียงซ้อน (Ensemble Model: EM) 53
3.4 การประเมินประสิทธิภาพของแบบจำลอง 54

บทที่ 4 ผลการวิจัยและอภิปรายผล 55

4.1 ผลลัพธ์ของแบบจำลองอัตราการว่างงาน 56
4.1.1 แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model) 59
4.1.1.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 59
4.1.1.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 62
4.1.2 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree Regression Model) 65
4.1.2.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 65

Ref. code: 25646304010116LCN

4.1.2.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 69
4.1.3 แบบจำลองซับพอรตเวกเตอร์รีเกรสชัน (Support Vector Regression Model: SVR) 72
4.1.3.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 73
4.1.3.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 73
4.1.4 แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN) 74
4.1.4.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 75
4.1.4.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 75
4.1.5 แบบจำลองแบบเรียงซ้อน (Ensemble Model: EM) 76
4.1.5.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 77
4.1.5.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 77
4.2 ผลการทดสอบแบบจำลองกับชุดข้อมูลทดสอบ (Testing Data) 78
4.2.1 รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) 79
4.2.2 Scatter plot 80
4.2.2.1 Linear Regression Model 81
4.2.2.2 Optimizable Decision Tree Regression Model 82
4.2.2.3 Optimizable Support Vector Regression Model 83
4.2.2.4 Neural Network Model 84
4.2.2.5 Optimizable Ensemble Model 85
4.2.3 ผลการเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ 86
4.2.3.1 Linear Regression Model 87
4.2.3.2 Optimizable Decision Tree Regression Model 88
4.2.3.3 Optimizable Support Vector Regression Model 89
4.2.3.4 Neural Network Model 90
4.2.3.5 Optimizable Ensemble Model 91

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 93

Ref. code: 25646304010116LCN

5.1 การสรุปและอภิปรายผลการศึกษา 93
5.2 ข้อจำกัดของงานวิจัย และข้อเสนอแนะ 95
5.3 การศึกษาต่อยอดในอนาคต 96

รายการอ้างอิง 97

Ref. code: 25646304010116LCN

สารบัญตาราง

ตารางที่		หน้า
2.1	การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ	28
2.2	การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวแปรอื่น ๆ	34
2.3	การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องที่ประยุกต์ใช้แบบจำลอง Machine Learning	37
4.1	กำหนดตัวแปร Google Trends ที่ใช้ในการศึกษา	57
4.2	ค่าสถิติเบื้องต้นของตัวแปรที่ศึกษา (Descriptive Statistics) ในชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธี Linear Regression	59
4.3	รายละเอียดแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)	61
4.4	สรุปค่าสถิติเบื้องต้นของตัวแปรที่ศึกษา (Descriptive Statistics) ในชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธี Linear Regression	62
4.5	รายละเอียดแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)	64
4.6	ค่าความคาดเคลื่อนรากที่สองกำลังสองเฉลี่ย (RMSE) ในแต่ละแบบจำลอง	79
4.7	ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Linear Regression	81
4.8	ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Linear Regression	81
4.9	ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Optimizable Tree	82
4.10	ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Optimizable Tree	82
4.11	ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Optimizable SVR	83
4.12	ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Optimizable SVR	83
4.13	ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Neural Network	84
4.14	ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Neural Network	84
4.15	ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Optimizable Ensemble	85

Ref. code: 25646304010116LCN

4.16 ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Optimizable Ensemble 85
5.1 สรุปผลประสิทธิภาพของแบบจำลอง 94

Ref. code: 25646304010116LCN

สารบัญภาพ

ภาพที่		หน้า
1.1	อัตราการว่างงานรายเดือนในประเทศไทยตั้งแต่ปี พ.ศ. 2554 – 2563	3
1.2	อัตราการว่างงานของผู้ประกันตนที่รับประโยชน์ทดแทนกรณีการว่างงานรายเดือนของสำนักงานประกันสังคมตั้งแต่ มกราคม พ.ศ. 2554 – ตุลาคม พ.ศ. 2564	4
1.3	ปริมาณคำค้นหา Google search words ที่เกี่ยวข้องกับคำค้นหาการหางานในประเทศไทยตั้งแต่ปี พ.ศ. 2554 – 2564	4
2.1	การแบ่งข้อมูลของ Decision Tree Regression	20
2.2	การแบ่งขอบเขตของ Support Vector Regression	22
2.3	องค์ประกอบของ Neural Network	25
2.4	การแบ่งแบบจำลองของ Ensemble Regression	26
3.1	อัตราการว่างงานรายเดือนในประเทศไทยตั้งแต่ปี พ.ศ. 2554 – 2563 โดยสำนักงานสถิติแห่งชาติ	42
3.2	อัตราการว่างงานของผู้ประกันตนที่รับประโยชน์ทดแทนกรณีการว่างงานรายเดือนของสำนักงานประกันสังคมตั้งแต่ มกราคม พ.ศ. 2554 – ตุลาคม พ.ศ. 2564	43
3.3	การแบ่งข้อมูลออกเป็น Training Data และ Testing Data	44
3.4	กรอบแนวคิดในการศึกษา	45
3.5	Support vector machine	51
3.6	โครงข่ายประสาทเทียมแบบหลายชั้น	52
3.7	ขั้นตอนการรวมแบบจำลองการพยากรณ์ของ Ensemble	53
4.1	การแบ่งข้อมูลของชุดข้อมูลการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (NSO)	57
4.2	การแบ่งข้อมูลของชุดข้อมูลผู้ประกันตนในประกันสังคม (SSO)	57
4.3	ค่าพารามิเตอร์กับแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานจากสำนักงานสถิติแห่งชาติ (NSO)	61
4.4	ค่าพารามิเตอร์กับแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)	64

Ref. code: 25646304010116LCN

4.5 แผนภูมิต้นไม้ตัดสินใจของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 66
4.6 ค่าแบบจำลองของฟังก์ชันวัตถุประสงค์จากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 68
4.7 ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 68
4.8 แผนภูมิต้นไม้ตัดสินใจของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 69
4.9 ค่าแบบจำลองของฟังก์ชันวัตถุประสงค์จากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 71
4.10 ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 72
4.11 ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable SVR ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 73
4.12 ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable SVR ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 74
4.13 โครงสร้างประสาทเทียมของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 75
4.14 โครงสร้างประสาทเทียมของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 76
4.15 ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Ensemble ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) 77
4.16 ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Ensemble ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) 78

Ref. code: 25646304010116LCN

4.17 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Linear Regression 87
4.18 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Linear Regression 87
4.19 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Decision Tree Regression 88
4.20 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Decision Tree Regression 88
4.21 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Support Vector Regression 89
4.22 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Support Vector Regression 89
4.23 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Neural Network 90
4.24 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Neural Network 90
4.25 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าาพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Ensemble 91
4.26 การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Ensemble 91

Ref. code: 25646304010116LCN

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในสถานการณ์ปัจจุบันหลาย ๆ ประเทศทั่วโลกต้องเผชิญกับภาวะเศรษฐกิจที่เปลี่ยนแปลงอย่างรวดเร็ว ทั้งข้อมูลข่าวสารต่าง ๆ ที่ส่งผลกระทบต่อเศรษฐกิจและพฤติกรรมของมนุษย์ อีกทั้งยังต้องเผชิญกับการแพร่ระบาดของไวรัสโคโรนา 2019 หรือโรคโควิด-19 (COVID-19) ผลกระทบดังกล่าว ส่งผลให้การดำเนินกิจกรรมทางเศรษฐกิจหยุดชะงัก ซึ่งจะส่งผลต่อการจ้างงานในตลาดแรงงาน และส่งผลให้อุตสาหกรรมได้รับผลกระทบจนต้องเลิกจ้างแรงงานเพื่อให้ธุรกิจสามารถลดต้นทุนได้ อีกทั้งยังส่งผลกระทบต่อระบบเศรษฐกิจเป็นวงกว้าง

อัตราการว่างงาน (Unemployment Rate) เป็นหนึ่งในตัวชี้วัดที่สำคัญในการแสดงถึงสถานการณ์เศรษฐกิจภายในประเทศ เนื่องจากสามารถบ่งบอกถึงรายได้ในครัวเรือนที่ส่งผลต่อการบริโภคภายในประเทศ โดยในประเทศไทยมีอัตราการว่างงานเฉลี่ยระหว่างปี พ.ศ. 2554 - 2562 อยู่ที่ร้อยละ 0.89 ขณะที่ในปี พ.ศ. 2563 ได้เกิดสถานการณ์การแพร่ระบาดของโรคโควิค-19 ทำให้มีอัตราการว่างงานเฉลี่ยระหว่างปีเพิ่มขึ้นเป็นร้อยละ 1.62 ซึ่งจะส่งผลต่ออัตราการเติบโตของผลิตภัณฑ์มวลรวมภายในประเทศที่แท้จริง (Real Gross Domestic Product Growth หรือ Real GDP Growth) ลดลง และเมื่อเปรียบเทียบกับข้อมูลจริงตามรายงานของ Federal Reserve Bank พบว่า ในปี พ.ศ. 2563 ประเทศไทยมีอัตราการเติบโตของ Real GDP ลดลงร้อยละ 6.09 จะเห็นได้ว่า อัตราการว่างงานเป็นตัวแปรที่มีลักษณะเป็นตัวชี้วัดนำ (Leading Indicator) โดยสามารถบ่งชี้ได้ว่าเศรษฐกิจจะชะลอตัวลงหรือไม่ เนื่องจากรายจ่ายด้านแรงงานเป็นรายจ่ายที่มีสัดส่วนมากในแต่ละธุรกิจ ทำให้นายจ้างมักจะมีการปรับเปลี่ยนอัตราการจ้างงานอย่างรวดเร็วในกรณีที่ต้องควบคุมต้นทุน อย่างไรก็ตาม อัตราการว่างงานก็เป็นตัวแปรที่มีค่าที่ตอบสนองช้าต่อการเปลี่ยนแปลงในเศรษฐกิจ (Lagging Indicator) เช่น บางครั้งตัวเลขการว่างงานจะยังสูงแม้ภาวะเศรษฐกิจถดถอยจะผ่านไปแล้ว เนื่องจากนายจ้างส่วนใหญ่ยังลังเลที่จะจ้างแรงงานเพิ่มจนกว่าจะแน่ใจว่าเศรษฐกิจกลับมาแข็งแรงและพร้อมที่จะเติบโตต่อไป นอกจากนี้ ในช่วงสถานการณ์โควิด-19 ยังพบว่า แรงงานไทยที่ว่างงานเป็นจำนวนมากอยู่นอกระบบและไม่รวมอยู่ในข้อมูลของอัตราการว่างงาน (Nikkei Asia, 2021) ซึ่งปัจจัยเหล่านี้ทำให้อัตราการว่างงานไทยที่ถูกประกาศอย่างเป็นทางการในปัจจุบันอยู่ในระดับต่ำกว่าความเป็นจริง ดังนั้น ผู้ศึกษาจึงมีความสนใจในการลดปัญหาข้อมูลของอัตราการว่างงานที่มีความล่าช้าและอยู่ใน

Ref. code: 25646304010116LCN

ระดับต่ำกว่าความเป็นจริงโดยการนำข้อมูลคำค้นหาที่เกี่ยวข้องกับการหางานและการว่างงานใน Google Trends มาประยุกต์ใช้

ข้อมูล Google Trends เป็นข้อมูลที่สะท้อนถึงความสนใจของประชากรในหลากหลายด้าน และยังเป็นข้อมูลที่สามารถแสดงถึงพฤติกรรมของผู้บริโภคที่มีความซับซ้อนและเปลี่ยนแปลงไปอย่างรวดเร็วได้ ซึ่งเป็นอีกหนึ่งทางเลือกที่มีศักยภาพในการติดตามภาวะและคาดการณ์แนวโน้มเศรษฐกิจได้ค่อนข้างทันที อีกทั้งยังเป็นข้อมูลที่มีความผูกพันใกล้ชิดของพฤติกรรมประชาชนกับเทคโนโลยีที่นับวันจะมีเพิ่มขึ้น เช่น อินเทอร์เน็ต รวมถึงความก้าวหน้าของโปรแกรมที่ช่วยจัดการกับข้อมูลขนาดใหญ่จะเป็นปัจจัยสำคัญที่ช่วยให้มีการใช้ประโยชน์จากแหล่งข้อมูลขนาดใหญ่ (Big Data) ในกระบวนการวิเคราะห์ ตัดสินใจ และดำเนินนโยบายของทั้งภาครัฐและภาคเอกชนเพิ่มขึ้น (ธนาคารแห่งประเทศไทย, 2561) นอกจากนี้ จากผลการสำรวจของกระทรวงดิจิทัล พบว่า จำนวนผู้ใช้งานอินเทอร์เน็ตเพิ่มขึ้นทุกปี โดยในปี พ.ศ. 2563 มีจำนวนผู้ใช้งานอินเทอร์เน็ตเพิ่มขึ้นเป็น 50.1 ล้านคน เพิ่มขึ้นจากปี พ.ศ. 2562 ร้อยละ 5.47 และเพิ่มขึ้นจากปี พ.ศ. 2558 ร้อยละ 26.84 และยังพบว่า ในปี พ.ศ. 2563 ผู้ที่ใช้งานอินเทอร์เน็ตใช้บริการอินเทอร์เน็ตเฉลี่ยวันละ 11 ชั่วโมง 25 นาที เพิ่มขึ้นจากปี พ.ศ. 2562 จำนวน 1 ชั่วโมง 3 นาที เนื่องจากในยุคปัจจุบันสามารถเข้าถึงอินเทอร์เน็ตได้ง่ายขึ้น มีเครือข่ายที่ครอบคลุม อีกทั้งสถานการณ์จากโรคโควิด-19 ส่งผลให้การใช้ชีวิตประจำวันมีความจำเป็นต้องใช้อินเทอร์เน็ตมากขึ้นเพื่อลดความเสี่ยงจากการต้องพบปะผู้คน ขณะที่เมื่อจำแนกตามกลุ่มอายุที่ใช้บริการอินเทอร์เน็ต พบว่า กลุ่มอายุ 20 – 39 ปี มีการใช้อินเทอร์เน็ตมากที่สุดซึ่งเป็นกลุ่มคนในวัยแรงงาน

ทั้งนี้ ข้อมูลจาก Google Trends ยังสามารถนำมาใช้ประโยชน์โดยการทำ Nowcasting หรือการคาดการณ์ปัจจุบัน โดยการใช้ข้อมูลที่มีอยู่ในช่วงเวลาปัจจุบันมาใช้คาดการณ์ในช่วงเวลาเดียวกัน เช่น การทำ GDP Nowcasting มาอธิบายตัวเลขทางเศรษฐกิจที่สำคัญ ซึ่งโดยปกติแล้วตัวเลขของ GDP จะเป็นตัวเลขที่มาจากกิจกรรมทางเศรษฐกิจ เช่น ตัวเลขด้านการผลิต ตัวเลขด้านตลาดแรงงาน ตัวเลขด้านการค้าต่างประเทศ ซึ่งตัวเลขเหล่านี้มักจะมีข้อมูลที่มีความถี่เป็นรายเดือนแล้วรวบรวมมาเป็นตัวเลขของ GDP ที่มีความถี่เป็นรายไตรมาส ซึ่งมีความล่าช้าเนื่องจากต้องใช้เวลาในการรวบรวมข้อมูล ส่งผลทำให้การคำนวณค่า GDP ในแต่ละไตรมาสต้องใช้เวลาประมาณ 1 - 2 เดือน แต่หากนำข้อมูลจาก Google Trends ที่มีความถี่เป็นรายสัปดาห์หรือรายเดือนจะสามารถคาดการณ์ GDP ในไตรมาสปัจจุบันได้

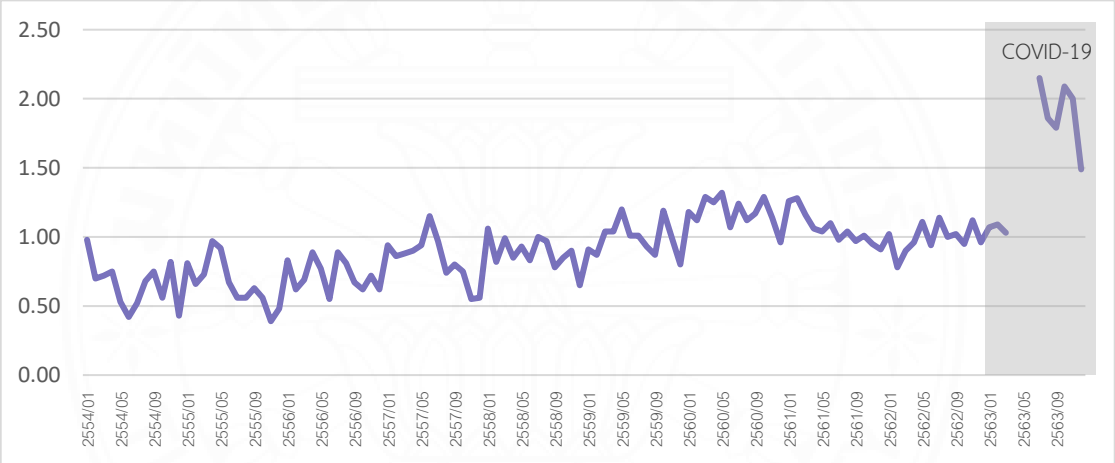
ขณะเดียวกันในปัจจุบัน ธนาคารกลางสหรัฐอเมริกาสาขาแอตแลนต้า (Federal Reserve Bank of Atlanta) ได้นำเสนอ Nowcasting Model ของ GDP เรียกว่า GDPNow โดยใช้หลักการของ Nowcasting ตามที่อธิบายข้างต้นเพื่อประยุกต์ในการคาดการณ์ GDP ในไตรมาส

Ref. code: 25646304010116LCN

ปัจจุบัน และจะทำการปรับเปลี่ยนให้เป็นข้อมูลแบบทันการณ์ เมื่อมีข้อมูลตัวเลขทางเศรษฐกิจต่าง ๆ ที่เกี่ยวข้องทยอยประกาศออกมา เช่นเดียวกันกับธนาคารกลางสหรัฐอเมริกาสาขานิวยอร์ค (Federal Reserve Bank of New York) ได้นำเสนอ Nowcasting Model ของ GDP เรียกว่า FRBNY Nowcast (วรประภา นาควัชระ, 2559) ดังนั้น วิธีการใช้ข้อมูลจาก Google Trends จึงเป็นที่น่าสนใจสำหรับการใช้เป็นดัชนีชี้วัดทางเศรษฐกิจต่าง ๆ โดยเฉพาะด้านอัตราการว่างงานที่เป็นปัญหาภายใต้สถานการณ์โควิด-19 ในปัจจุบัน

ภาพที่ 1.1

อัตราการว่างงานรายเดือนในประเทศไทยตั้งแต่ปี พ.ศ. 2554 – 2563



*เดือนเมษายนถึงมิถุนายน ปี 2563 ไม่พบข้อมูลอัตราการว่างงาน

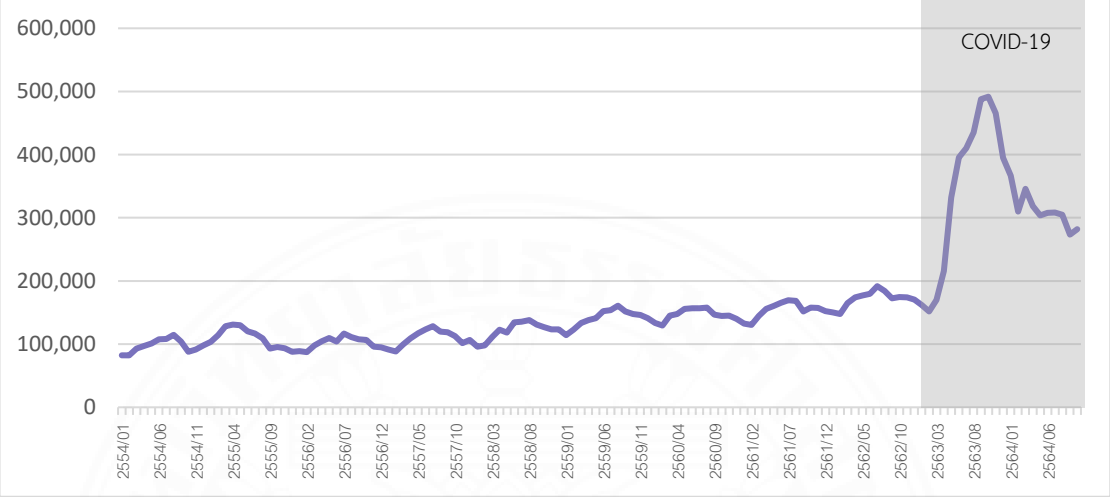
หมายเหตุ. โดยสำนักงานสถิติแห่งชาติ, 2564, (https://www.bot.or.th/App/BIZSHR/stat/Graph/)

จากภาพที่ 1.1 แสดงถึงอัตราการว่างงานรายเดือนในประเทศไทย พบว่ามีลักษณะเป็นวัฏจักรรายเดือนหรือมีการเปลี่ยนแปลงตามฤดูกาล โดยเมื่อไม่รวมข้อมูลอัตราการว่างงานในปี พ.ศ. 2563 เนื่องจากไม่มีการเก็บข้อมูลในเดือนเมษายน – มิถุนายน พบว่า จะมีอัตราการว่างงานน้อยที่สุดในเดือนธันวาคม ขณะที่มีอัตราการว่างงานมากที่สุดในเดือนมกราคม อย่างไรก็ตาม จากข้อมูลของสำนักงานสภาพัฒนาเศรษฐกิจและสังคมแห่งชาติในช่วงวิกฤตโควิด-19 พบว่า ผู้จบการศึกษาระดับอาชีวศึกษาและอุดมศึกษามีการว่างงานเพิ่มขึ้น โดยผู้ที่จบการศึกษาระดับอุดมศึกษามีอัตราการว่างงานร้อยละ 3.44 สะท้อนให้เห็นว่า การว่างงานในปัจจุบันอยู่ในกลุ่มแรงงานทักษะสูง (TNN, 2564)

Ref. code: 25646304010116LCN

ภาพที่ 1.2

อัตราการว่างงานของผู้ประกันตนที่รับประโยชน์ทดแทนกรณีการว่างงานรายเดือนของสำนักงานประกันสังคมตั้งแต่ มกราคม พ.ศ. 2554 – ตุลาคม พ.ศ. 2564

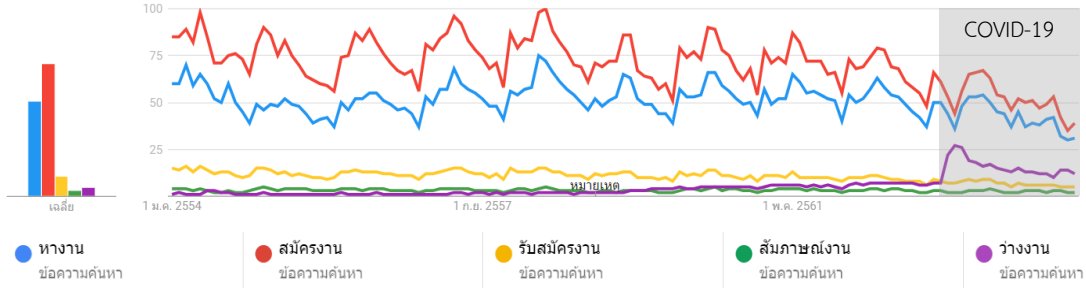


หมายเหตุ. โดยสำนักงานประกันสังคม, 2564, (http://warning.mol.go.th/uploadFile/pdf/pdf-2021-11-24-1637721494.pdf)

นอกจากนี้ ยังมีข้อมูลการว่างงานจากสำนักงานประกันสังคม ที่เป็นการเก็บข้อมูลในระบบผู้ประกันตนที่ขอรับประโยชน์ทดแทนกรณีการว่างงาน โดยที่ผู้ว่างงานในระบบ คือ ผู้ขอรับประโยชน์ทดแทนจากการว่างงาน และเมื่อรับสิทธิประโยชน์ครบกำหนดจะไม่ถูกนับว่าเป็นผู้ว่างงานในระบบอีกต่อไป ดังภาพที่ 1.2 พบว่ามีลักษณะเป็นวัฏจักรรายเดือนหรือมีการเปลี่ยนแปลงตามฤดูกาลเช่นเดียวกัน

ภาพที่ 1.3

ปริมาณคำค้นหา Google search words ที่เกี่ยวข้องกับคำค้นหาการหางานในประเทศไทยตั้งแต่ปี พ.ศ. 2554 – 2564



หมายเหตุ. โดย Google Trends, 2564, (https://trends.google.co.th/trends/?geo=TH)

จากภาพที่ 1.3 แสดงถึงคำค้นหาที่เกี่ยวข้องกับการหางาน โดยพบว่ามีลักษณะเปลี่ยนแปลงไม่คงที่ตามช่วงเวลาที่ต่างกัน ผู้ศึกษาจึงมีความสนใจที่จะนำข้อมูลนี้ไปหาความสัมพันธ์กับอัตราการว่างงานในประเทศไทย และสร้างแบบจำลองประมาณการอัตราการว่างงานในประเทศไทย ซึ่งจะนำผลไปใช้ในการสร้างดัชนีชี้นำเศรษฐกิจ

จากที่กล่าวมาข้างต้น ผู้ศึกษาจึงเห็นควรว่าหากมีการสร้างสมการ หรือแบบจำลองที่สามารถพยากรณ์อัตราการว่างงานที่สามารถใช้ Google Trends เป็นตัวแปรที่ช่วยพยากรณ์ได้อย่างเป็นปัจจุบันมากขึ้น โดยใช้วิธีการสมัยใหม่ที่มีความทันสมัย หลากหลายสามารถเลือกใช้ได้เหมาะสมกับลักษณะของข้อมูล และมีความแม่นยำ กล่าวคือ การใช้อัลกอริทึมของปัญญาประดิษฐ์ (Machine Learning) มาใช้เพื่อพัฒนาแบบจำลอง ซึ่งเป็นการต่อยอดจากผู้ที่เคยศึกษาไว้ และเพิ่มประสิทธิภาพแบบจำลองให้มีความแม่นยำมากยิ่งขึ้น

ด้วยเหตุผลนี้ ผู้ศึกษาจึงเห็นถึงความสำคัญของอัตราการว่างงานในประเทศไทยที่ได้รับผลกระทบจากสถานการณ์การแพร่ระบาดของโรคโควิด-19 ทั้งจากข้อมูลที่มีความล่าช้าและต่ำกว่าอัตราการว่างงานที่แท้จริง เนื่องจากมีแรงงานไทยที่อยู่นอกระบบอีกเป็นจำนวนมาก อีกทั้งยังมีลักษณะการเปลี่ยนแปลงที่ไม่คงที่ตามช่วงเวลาที่ต่างกันซึ่งสอดคล้องกับข้อมูลจาก Google Trends จึงมีความสนใจที่จะศึกษาความสัมพันธ์ระหว่างคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends กับอัตราการว่างงานในประเทศไทย ตลอดจนการนำคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends ไปประมาณการอัตราการว่างงานในอนาคต ถึงแม้ว่าจะมีผู้ศึกษาในลักษณะนี้บ้างแล้ว แต่ในการศึกษาครั้งนี้จะใช้ช่วงเวลาที่ประชากรไทยใช้บริการอินเทอร์เน็ตมากกว่าการศึกษาในอดีต และยังเป็นช่วงที่ได้รับผลกระทบจากโควิด-19 ที่ส่งผลกระทบต่ออัตราการว่างงานทำให้อัตราการว่างงานเพิ่มสูงขึ้น ผู้ศึกษาคาดว่าผลการศึกษาในครั้งนี้จะประมาณการอัตราการว่างงานในอนาคตได้ สามารถนำข้อมูลที่ได้ไปดำเนินนโยบายเพื่อให้บรรลุเป้าหมายทางเศรษฐกิจ

1.2 วัตถุประสงค์ของการศึกษา

1) เพื่อใช้ข้อมูลคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends มาประมาณการอัตราการว่างงานในอนาคต โดยการสร้างแบบจำลองด้วยวิธีการของ Machine Learning ที่มีความถูกต้อง และแม่นยำ

2) เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการใช้ข้อมูลคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends มาประมาณการอัตราการว่างงานในอนาคต ในแต่ละวิธีการของ Machine Learning

Ref. code: 25646304010116LCN

1.3 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อทราบถึงวิธีการสร้างแบบจำลองด้วยการใช้อัลกอริทึมแบบต่าง ๆ ของ Machine Learning ระหว่างคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends กับอัตราการว่างงานในประเทศไทย และสามารถนำข้อมูลที่ได้ไปประมาณการอัตราการว่างงานในอนาคต ตลอดจนทราบประสิทธิภาพของแบบจำลองที่ใช้ Machine Learning ในแต่ละวิธีการที่แตกต่างกัน อีกทั้งยังสามารถลดความล่าช้าในการรวบรวมข้อมูลอัตราการว่างงานจากภาครัฐ และสามารถนำข้อมูลที่ได้ไปใช้ประมาณการอัตราการว่างงานในอนาคตก่อนที่จะมีการรวบรวมข้อมูลที่แท้จริง เพื่อช่วยให้ภาครัฐสามารถใช้นโยบายทางเศรษฐกิจได้อย่างเหมาะสมและภาคเอกชนสามารถปรับตัวกับภาวะเศรษฐกิจที่กำลังจะเกิดขึ้นในอนาคต

1.4 ขั้นตอนการดำเนินงานวิจัย

1) ศึกษาบทความ วรรณกรรมปริทัศน์ และงานวิจัยที่เกี่ยวข้อง เพื่อรวบรวมข้อมูลที่ใช้ในการสร้างแบบจำลองการพยากรณ์อัตราการว่างงานด้วยคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends ในประเทศไทย ตลอดจนศึกษาวิธีการในการพยากรณ์ด้วยอัลกอริทึมของ Machine Learning หลากหลายโปรแกรม เพื่อให้ได้แบบจำลองที่มีวิธีการทันสมัย และมีความแม่นยำ

2) ศึกษารวบรวมคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends ที่เป็นปัจจัยที่มีความสำคัญ และส่งผลต่อการพยากรณ์อัตราการว่างงาน

3) เก็บรวบรวมข้อมูลตัวแปรที่ใช้ในการสร้างแบบจำลอง และข้อมูลวิธีการสร้างแบบจำลองในรูปแบบต่าง ๆ ได้แก่ Linear Regression Model, Optimizable Tree Regression Model, Optimizable Support Vector Regression Model, Neural Network Model และ Optimizable Ensemble Model เพื่อนำมาสร้างแบบจำลองการพยากรณ์อัตราการว่างงาน

4) เปรียบเทียบประสิทธิภาพและวิเคราะห์การพยากรณ์จากแบบจำลองที่สร้างจากอัลกอริทึมของ Machine Learning ด้วยวิธี Root Mean Square Error (RMSE)

1.5 ขอบเขตการศึกษา

ศึกษาสร้างแบบจำลองการพยากรณ์อัตราการว่างงานด้วยคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends ในประเทศไทย ด้วยอัลกอริทึมของ Machine Learning หลากหลาย

Ref. code: 25646304010116LCN

วิธีการ ได้แก่ Linear Regression Model, Optimizable Tree Regression Model, Optimizable Support Vector Regression Model, Neural Network Model และ Optimizable Ensemble Model โดยใช้ข้อมูลที่ใช้ในการสร้างแบบจำลองเป็นลักษณะอนุกรมเวลา (Time Series) แบบรายเดือน ซึ่งประกอบไปด้วยข้อมูล 2 ชุดข้อมูล ได้แก่ ข้อมูลอัตราการว่างงานจากการสำรวจประชากรของสำนักงานสถิติแห่งชาติ ซึ่งจะใช้ระยะเวลาของข้อมูลในการศึกษา 10 ปี ย้อนหลัง (ปี พ.ศ. 2554 – 2563) และอัตราการว่างงานจากผู้ประกันตนที่ขอรับประโยชน์ทดแทนในกรณีว่างงาน ซึ่งจะใช้ระยะเวลาของข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2554 – ตุลาคม พ.ศ. 2564 จากนั้นจะนำคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends ไปประมาณการอัตราการว่างงาน ซึ่งมีขอบเขตในการศึกษา ดังนี้

1.5.1 ข้อมูลของคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends

ข้อมูลที่ใช้จะมีค่าอยู่ระหว่าง 0 – 100 โดยเป็นการบ่งบอกถึงระดับความสนใจของประชาชนหรือความสนใจค้นหาทาง Google ในแต่ละช่วงเวลา และมีความถี่เป็นรายเดือน ส่วนคำค้นหาที่เกี่ยวข้อง ประกอบด้วย

กลุ่มที่ 1 สมัครงาน, หางาน, ว่างงาน, ประกันสังคม, และสัมภาษณ์งาน ซึ่งเป็นคำที่มีค่าสหสัมพันธ์ (correlation) กับอัตราการว่างงานเท่ากับ 0.3 ขึ้นไป โดยมีที่มาจากงานวิจัยของ Voraprapa & Nuarpear (2018) ได้ศึกษาเรื่อง Predicting the Present Revisited: The Case of Thailand และคำว่า สัมภาษณ์งาน, ประกันสังคม และว่างงาน เป็นคำที่มีค่าสหสัมพันธ์ (correlation) กับอัตราการว่างงานมากกว่าหรือเท่ากับ 0.3 โดยมีที่มาจากงานวิจัยของ ธนัตถ์พราวลีลารัตน์ (2562) ได้ศึกษาเรื่อง การเปรียบเทียบความสัมพันธ์ระหว่างอัตราการว่างงานไทยกับข้อมูลที่เกี่ยวข้องบนอินเทอร์เน็ตด้วยกูเกิลเทรนดส์ (Google Trends) และทวิตเตอร์ (Twitter)

กลุ่มที่ 2 นายจ้างหาคนงาน, งานราชการ, สมัครงานบริษัท, หางานสมัครงาน, ใบสมัครงาน, หางานทำ, สิทธิประกันสังคม และตกงาน เป็นต้น จากคำที่กล่าวมาข้างต้นคำว่า สมัคร

Ref. code: 25646304010116LCN

งาน, หางาน, ว่างงาน และประกันสังคม มีที่มาจากคำค้นหาที่เกี่ยวข้องที่มีลักษณะเป็นคำยอดนิยม[1] และคำมาแรง[2] ใน Google Trends

กลุ่มที่ 3 Jobs DB, Job Top Gun, LinkedIn, Job Thai, Job TH, กรมการจัดหางาน และ Job Fair มีที่มาจากชื่อบริษัทจัดหางาน

กลุ่มที่ 4 เงินชดเชยว่างงาน, ลงทะเบียนว่างงาน, กองทุนสำรองเลี้ยงชีพ และเงินชดเชยเลิกจ้าง มีที่มาจากคำที่เกี่ยวข้องกับสวัสดิการแรงงาน

1.5.2 ข้อมูลอัตราการว่างงาน

สำหรับข้อมูลอัตราการว่างงานจะประกอบไปด้วยข้อมูล 2 ชุด ได้แก่

(1) ข้อมูลอัตราการว่างงานรายเดือนจากเว็บไซต์ของธนาคารแห่งประเทศไทย โดยข้อมูลที่นำมาใช้จะมาจากการสำรวจประชากรของสำนักงานสถิติแห่งชาติ (National Statistical Office: NSO)

(2) ข้อมูลอัตราการว่างงานรายเดือนจากสำนักงานประกันสังคม ซึ่งเป็นข้อมูลของผู้ประกันตนในประกันสังคมที่ขอรับผลประโยชน์ทดแทนในกรณีว่างงาน (Social Security Office: SSO)

---

[1] คำยอดนิยม คือ คำค้นหาที่ได้รับความนิยมสูงสุด การให้คะแนนจะเป็นระดับที่สัมพัทธ์กัน โดยที่ 100 คือคำค้นหาที่ค้นหามากที่สุด, ค่า 50 คือคำค้นหาที่มีการค้นหาบ่อยเป็นครึ่งหนึ่งของคำค้นหาที่ได้รับความนิยมสูงสุด และลดหลั่นลงมาเรื่อย ๆ

[2] คำมาแรง - คำค้นหาที่มีจำนวนการค้นหาเพิ่มขึ้นมากที่สุดนับตั้งแต่ระยะเวลาล่าสุด ผลลัพธ์ที่ระบุว่า "ดาวรุ่งพุ่งแรง" แสดงการเพิ่มขึ้นอย่างมาก ซึ่งอาจเป็นเพราะคำค้นหาเหล่านี้เป็นคำใหม่และมีการค้นหาก่อนหน้า (หากมี) เพียงเล็กน้อย

Ref. code: 25646304010116LCN

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎีที่เกี่ยวข้องกับการมีงานทำและการว่างงาน

2.1.1 นิยามของตัวแปรที่ใช้ในการศึกษา

ประกอบไปด้วย 3 ตัวแปร ดังนี้

2.1.1.1 อัตราการว่างงานจากการสำรวจประชากร

จากการทบทวนวรรณกรรมต่าง ๆ และแหล่งที่มาของคำนิยามของการว่างงานและการมีงานทำ พบว่า มักจะมีการปรับเปลี่ยนอย่างสม่ำเสมอ โดยมีวัตถุประสงค์เพื่อให้สอดคล้องกับสภาพทางสังคม และเศรษฐกิจปัจจุบัน รวมไปถึงเพื่อเป็นประโยชน์ต่อผู้ที่ต้องการใช้ข้อมูล และยังต้องสอดคล้องกับมาตรฐานสากลขององค์การแรงงานระหว่างประเทศ (International Labor Organization: ILO) และองค์กรสหประชาชาติ (United Nations: UN) สำหรับนิยามของการทำงานและการว่างงานที่จะใช้ในงานวิจัยนี้จะใช้นิยามของสำนักงานสถิติแห่งชาติเพื่อให้สอดคล้องกับที่มาของตัวแปรอัตราการว่างงานที่ใช้ศึกษา โดยสำนักงานสถิติแห่งชาติจะทำการสำรวจภาวการณ์ทำงาน (Labor Force Survey: LFS) ของประชากรทั่วประเทศไทย ซึ่งเริ่มใช้มาตั้งแต่ พ.ศ. 2526 และมีการปรับปรุงอย่างต่อเนื่อง จนกระทั่งปี พ.ศ. 2544 ได้มีการกำหนดอายุขั้นต่ำของประชากรวัยทำงานเป็น 15 ปี โดยจะมีนิยามที่เกี่ยวกับการมีงานทำและการว่างงานตามแบบสำรวจภาวะการทำงานของประชากร ดังนี้

(1) ผู้มีงานทำ หมายถึง บุคคลที่มีอายุ 15 ปีขึ้นไป และในการสำรวจต้องมีลักษณะอย่างใดอย่างหนึ่ง ดังต่อไปนี้

- ได้ทำงานตั้งแต่ 1 ชั่วโมงขึ้นไป โดยได้รับค่าจ้าง เงินเดือน ผลกำไร เงินปันผล ค่าตอบแทนที่มีลักษณะอย่างอื่นสำหรับผลงานที่ทำอาจเป็นเงินสดหรือสิ่งของ
- ไม่ได้ทำงาน หรือทำงานน้อยกว่า 1 ชั่วโมง แต่เป็นบุคคลที่มีลักษณะอย่างใดอย่างหนึ่งใน 2 กลุ่ม ดังนี้

<u>กลุ่มที่ 1</u> ยังได้รับค่าตอบแทน ค่าจ้าง ผลประโยชน์อื่น ๆ ผลกำไรจากงาน ธุรกิจในระหว่างที่ไม่ได้ทำงาน

<u>กลุ่มที่ 2</u> ไม่ได้ค่าผลตอบแทน ค่าจ้าง ผลประโยชน์อื่น ๆ ผลกำไรจากงาน ธุรกิจในระหว่างที่ไม่ได้ทำงานแต่ยังมีงานหรือธุรกิจที่จะกลับไปทำ

Ref. code: 25646304010116LCN

- ทำงานอย่างน้อย 1 ชั่วโมง โดยไม่ได้รับค่าจ้างในธุรกิจ หรือทำไร่นาเกษตรของหัวหน้าครัวเรือนหรือของสมาชิกในครัวเรือน

(2) ผู้ว่างงาน หมายถึง บุคคลที่มีอายุ 15 ปี ขึ้นไป และในการสำรวจจะมีลักษณะอย่างใดอย่างหนึ่ง ดังต่อไปนี้

- ไม่ได้ทำงานและไม่มีงานประจำ แต่ได้หางาน สมัครงานหรือรอการบรรจุ ในระหว่าง 30 วัน ก่อนวันสัมภาษณ์
- ไม่ได้ทำงานและไม่มีงานประจำ และไม่ได้หางานทำในระหว่าง 30 วัน ก่อนวันสัมภาษณ์ แต่พร้อมที่จะทำงานในช่วงที่มีการสำรวจแรงงาน

(3) กำลังแรงงานปัจจุบัน หมายถึง บุคคลที่มีอายุ 15 ปีขึ้นไป ซึ่งในช่วงการสำรวจมีงานทำหรือว่างงาน ตามนิยามในข้อ (1) หรือ (2)

(4) กำลังแรงงานที่รอฤดูกาล หมายถึง บุคคลที่มีอายุ 15 ปีขึ้นไป โดยในช่วงการสำรวจเป็นผู้ไม่เข้าข่ายคำนิยามของผู้มีงานทำ หรือผู้ว่างงาน แต่เป็นผู้รอฤดูกาลที่เหมาะสมเพื่อที่จะทำงาน และเป็นบุคคลที่ตามปกติจะทำงานที่ไม่ได้รับสิ่งตอบแทนในไร่นาเกษตร หรือธุรกิจซึ่งทำกิจกรรมตามฤดูกาล โดยมีหัวหน้าครัวเรือน หรือสมาชิกคนอื่น ๆ ในครัวเรือนเป็นเจ้าของหรือผู้ดำเนินการ

(5) กำลังแรงงานรวม หมายถึง บุคคลทุกคนที่มีอายุ 15 ปีขึ้นไป ในช่วงที่สำรวจภาวะการทำงานเป็นผู้ที่อยู่ในกำลังแรงงานปัจจุบัน หรือเป็นผู้ถูกจัดจำแนกอยู่ในประเภทกำลังแรงงานที่รอฤดูกาลตามคำนิยามที่ได้ระบุไว้

(6) ผู้ไม่อยู่ในกำลังแรงงาน หมายถึง บุคคลที่ไม่เข้าข่ายคำนิยามของผู้อยู่ในกำลังแรงงานในช่วงการสำรวจ โดยเป็นบุคคลซึ่งในช่วงการสำรวจมีอายุ 15 ปีขึ้นไป แต่ไม่ได้ทำงานและไม่พร้อมที่จะทำงานเนื่องจากเป็นผู้ที่

- ทำงานบ้าน
- เรียนหนังสือ
- ยังเด็กเกินไป หรือชนชรามาก
- ไม่สามารถทำงานได้เนื่องจากพิการทางร่างกายหรือจิตใจ หรือเจ็บป่วยเรื้อรัง
- พักผ่อน
- เกษียณการทำงาน
- ไม่พร้อมที่จะทำงานเนื่องจากเหตุผลอื่น เช่น ไม่สมัครใจทำงาน ทำงานโดยไม่ได้รับค่าจ้าง ผลกำไร ส่วนแบ่ง หรือสิ่งตอบแทนอื่น ๆ ให้แก่บุคคลซึ่งมิได้เป็นสมาชิกใน

Ref. code: 25646304010116LCN

ครัวเรือนเดียวกัน ทำงานให้แก่องค์การ หรือสถาบันการกุศลต่าง ๆ โดยไม่ได้รับค่าจ้างผลกำไรส่วนแบ่งหรือสิ่งตอบแทนอย่างใด

(7) งาน หมายถึง กิจการที่ทำที่มีลักษณะอย่างใดอย่างหนึ่ง ดังต่อไปนี้

- กิจการที่ทำแล้วได้รับค่าตอบแทนเป็นเงินหรือสิ่งของ ค่าตอบแทนที่เป็นเงิน อาจจ่ายเป็นรายเดือน รายสัปดาห์รายวัน หรือรายชิ้น
- กิจการที่ทำแล้วได้ผลกำไร หรือหวังที่จะได้รับผลกำไร หรือส่วนแบ่งเป็นการตอบแทน
- กิจการที่ทำให้กับธุรกิจของสมาชิกในครัวเรือน โดยไม่ได้รับค่าจ้างหรือผลกำไรตอบแทนอย่างใดซึ่งสมาชิกในครัวเรือนที่ประกอบธุรกิจนั้นจะมีสถานภาพการทำงานเป็นประกอบธุรกิจส่วนตัว หรือนายจ้าง

ส่วนการคำนวณอัตราการว่างงานจากการสำรวจประชากรของสำนักงานสถิติแห่งชาติ มีวิธีการคำนวณอัตราการว่างงาน ดังนี้

$$\frac{\text{ผู้ไม่มีงานทำ}}{\text{กำลังแรงงานรวม}} \times 100$$

สำหรับประเภทของการว่างงานตามเศรษฐศาสตร์ วรลักษณ์ หิมะกลัส (2555, น. 72 - 73) ได้ให้ความหมายของการว่างงานไว้ว่า “ภาวการณ์ที่ผู้ที่อยู่ในวัยทำงานซึ่งมีความสมัครใจและมีความสามารถที่จะทำงาน ณ ระดับค่าแรงที่ปรากฏ แต่ไม่สามารถหางานทำได้” และได้แบ่งประเภทของการว่างงาน เป็น 2 ประเภท ดังนี้

(1) การว่างงานโดยเปิดเผย (Open Unemployment) คือ การว่างงานในระหว่างเวลาที่ทำการสำรวจ สามารถแบ่งได้เป็น

- **การว่างงานชั่วคราว (Frictional Unemployment)** คือ บุคคลที่ไม่ได้ทำงานหรือว่างงานเพียงชั่วระยะเวลาสั้น ๆ
- **การว่างงานตามฤดูกาล (Seasonal Unemployment)** คือ การว่างงานที่เกิดขึ้นเนื่องจากการผลิตสินค้าบางอย่างต้องทำการผลิตเป็นฤดูกาล
- **การว่างงานเนื่องจากโครงสร้างของระบบเศรษฐกิจ (Structural Unemployment)** คือ การว่างงานที่เกิดเนื่องจากการเปลี่ยนแปลงของสภาพแวดล้อมทางเศรษฐกิจ เช่น รสนิยมของผู้บริโภค เทคนิควิทยาการ การเพิ่มขึ้นของประชากร
- **การว่างงานเนื่องจากวัฏจักรเศรษฐกิจ (Cyclical Unemployment)** โดยวัฏจักรเศรษฐกิจ (Business Cycles) สามารถแบ่งได้เป็น 4 ระยะ ได้แก่ ระยะเศรษฐกิจรุ่งเรือง

Ref. code: 25646304010116LCN

(Prosperous period) ระยะเศรษฐกิจหดตัว (Recession period) ระยะเศรษฐกิจตกต่ำ (Depression period) ระยะเศรษฐกิจฟื้นตัว (Recovery period)

(2) การว่างงานแอบแฝง (Disguised Unemployment) คือ การมีจำนวนแรงงานเกินความจำเป็น ส่วนใหญ่มักพบในภาคเกษตรกรรม บุคคลที่ว่างงานแฝงจะทำงานต่ำว่าระดับ (Underemployment)

2.1.1.2 อัตราการว่างงานจากผู้ประกันตนในประกันสังคมที่ขอรับประโยชน์ทดแทนกรณีว่างงาน

เป็นผู้ประกันตนในฐานข้อมูลประกันสังคมที่ต้องการรับประโยชน์ทดแทนประกอบไปด้วย 2 กรณี ได้แก่ กรณีถูกเลิกจ้าง และกรณีลาออกหรือสิ้นสุดสัญญาจ้างตามกำหนดระยะเวลา ทั้งนี้ ต้องมีคุณลักษณะเป็นลูกจ้างซึ่งเป็นผู้ประกันตนจะต้องเป็นผู้ประกันตนที่มีสิทธิได้รับประโยชน์ทดแทนในกรณีว่างงาน ตามมาตรา 78 ตามพระราชบัญญัติประกันสังคม พ.ศ. 2533 กล่าวคือ ต้องมีการจ่ายเงินสมทบมาแล้วไม่น้อยกว่า 6 เดือน และต้องอยู่ภายในระยะเวลา 15 เดือนก่อนการว่างงาน (สำนักงานประกันสังคม, 2558) ซึ่งในงานศึกษานี้จะคำนวณอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) จากข้อมูลประกอบด้วย

(1) ผู้ประกันตนภาคบังคับตามมาตรา 33 คือ ลูกจ้างผู้ซึ่งทำงานให้กับนายจ้างที่อยู่ในสถานประกอบการที่มีลูกจ้างตั้งแต่ 1 คนขึ้นไป

(2) ผู้ประกันตนภาคสมัครใจตามมาตรา 39 คือ ผู้ประกันตนโดยสมัครใจเป็นบุคคลที่เคยเป็นผู้ประกันตนมาตรา 33 จ่ายเงินสมทบก่อนออกจากงานมาแล้วไม่น้อยกว่า12เดือนแล้วลาออก แต่ต้องการรักษาสิทธิประกันสังคม

(3) ผู้ประกันตนภาคสมัครใจตามมาตรา 40 คือ บุคคลทั่วไปที่ประกอบอาชีพอิสระหรือแรงงานนอกระบบไม่เป็นผู้ประกันตนมาตรา 33 หรือผู้ประกันตนมาตรา 39 (สำนักงานประกันสังคม, 2562)

โดยสามารถเขียนวิธีการคำนวณได้ ดังนี้

$$\text{อัตราการว่างงาน} = \frac{\text{ผู้ประกันตนที่รับประโยชน์ทดแทนกรณีว่างงาน}}{\text{ผู้ประกันตนตามมาตรา 33 มาตรา 39 และมาตรา 40}}$$

Ref. code: 25646304010116LCN

2.1.1.3 Google Trends

เป็นเครื่องมือที่สามารถใช้วิเคราะห์แนวโน้มและระดับความสนใจของข้อความต่าง ๆ ที่มีการค้นหาผ่านเว็บไซต์ Google เมื่อเทียบกับจุดสูงสุดในช่วงเวลาที่ระบุ ซึ่งมีค่าตั้งแต่ 0 ถึง 100 โดย 100 หมายถึง ระดับความสนใจสูงสุด

2.1.2 ทฤษฎีเกี่ยวกับการมีงานทำและการว่างงาน

ทฤษฎีการจ้างงานทางเศรษฐศาสตร์ จะอธิบายถึงระดับการจ้างงานที่เกิดจากอุปสงค์และอุปทานแรงงาน โดยหากอุปสงค์น้อยกว่าอุปทานแรงงานจะทำให้เกิดการว่างงาน ซึ่งระดับการจ้างงานนี้จะมีความสัมพันธ์กับระดับผลผลิตในทางเดียวกัน กล่าวคือ เมื่อระดับการจ้างงานเพิ่มขึ้นจะส่งผลต่อระดับผลผลิตที่สูงขึ้นด้วย ซึ่งจะมีหลายทฤษฎีที่สามารถอธิบายได้ ดังนี้

2.1.2.1 ทฤษฎีการว่างงานของสำนักคลาสสิค (The classical Theory of Employment)

เป็นทฤษฎีที่ให้ความสนใจในเรื่องของการจ้างงานเต็มที่ (Full Employment) โดยที่ ณ ระดับการจ้างงานเต็มที่ถ้ามีการว่างงานเกิดขึ้นจะเป็นการว่างงานโดยความสมัครใจ เนื่องจากมีแนวคิดที่ว่า การว่างงานที่เกิดขึ้น เกิดจากคนงานเรียกร้องค่าจ้างที่เป็นตัวเงินสูงเกินไปและไม่ยอมรับค่าจ้าง ณ ระดับค่าจ้างที่ปรากฏ จึงสมัครใจที่จะว่างงาน แต่ถ้าคนงานยอมรับค่าจ้างที่เป็นตัวเงินที่ต่ำลง ค่าจ้างที่แท้จริงจะลดลง ระดับการจ้างงานก็จะเพิ่มขึ้นถึงระดับการจ้างงานเต็มที่การว่างงานก็จะหมดไป และยังมีความเชื่อที่ว่า ระดับอัตราค่าจ้างที่เป็นตัวเงินจะถูกกำหนดโดยการต่อรองระหว่างผู้ประกอบการและแรงงาน และค่าจ้างที่เป็นตัวเงินจะเป็นปัจจัยกำหนดค่าจ้างที่แท้จริง แรงงานจึงสามารถกำหนดระดับค่าจ้างที่แท้จริง และกำหนดระดับการจ้างงานได้

ดังนั้น ณ ระดับค่าจ้างถ้ามีการว่างงานเกิดขึ้นจะเป็นการว่างงานโดยสมัครใจ นักเศรษฐศาสตร์สำนักคลาสสิคจึงมีความเห็นว่า อุปสงค์และอุปทานแรงงานเป็นปัจจัยกำหนดระดับการจ้างงาน อัตราค่าจ้างและอัตราค่าจ้างที่แท้จริงจะอยู่ ณ จุดตัดของเส้นอุปสงค์และอุปทานแรงงาน นอกจากนี้มีแนวคิดของสำนักคลาสสิคว่า ระบบเศรษฐกิจจะอยู่ ณ ระดับการจ้างงานเต็มที่เสมอ และเป็นระดับที่ผลผลิตสูงสุด อุปสงค์รวมของระบบเศรษฐกิจเท่ากับอุปทานรวมเสมอ การว่างงานอาจเกิดขึ้นได้แต่เพียงชั่วคราวและสามารถแก้ไขได้เองโดยไม่ต้องให้รัฐเข้ามาแทรกแซง

2.1.2.2 ทฤษฎีการว่างงานของสำนักเคนส์ (John Maynard Keynes)

เป็นทฤษฎีที่สามารถใช้ในระบบเศรษฐกิจที่มีการจ้างงานเต็มที่ และการว่างงานได้ โดยยอมรับทฤษฎีของสำนักคลาสสิคที่ว่าระดับผลผลิตมีความสัมพันธ์โดยตรงกับการจ้างงาน แต่มีข้อแย้งกับสำนักคลาสสิค กล่าวคือ สำนักเคนส์จะมองว่าผลผลิตจะถูกกำหนดโดยอุปสงค์รวมของระบบเศรษฐกิจซึ่งไม่จำเป็นจะต้องเท่ากับอุปทานรวมเสมอไป อุปสงค์รวมอาจต่ำกว่าอุปทาน

Ref. code: 25646304010116LCN

รวมได้ ดังนั้น ระดับการจ้างงานจึงอาจต่ำกว่าระดับการจ้างงานเต็มที่ นอกจากนี้ เคนส์ยังเชื่อว่า ถ้าอัตราค่าจ้างที่แท้จริงลดลงจะทำให้รายจ่ายเพื่อการบริโภคลดลง ทำให้หน่วยธุรกิจขายสินค้าที่ผลิตขึ้นได้ลดลง แสดงให้เห็นว่าการลดอัตราค่าจ้างจะทำให้การว่างงานรุนแรงยิ่งขึ้น และถ้าอัตราค่าจ้างที่แท้จริงลดลง ซึ่งอาจเป็นผลจากการลดลงของค่าจ้างที่เป็นตัวเงิน แต่ระดับราคาสินค้าคงที่หรือระดับราคาสินค้าเพิ่มขึ้นในอัตราที่ช้ากว่าอัตราเพิ่มค่าจ้างที่เป็นตัวเงิน จะส่งผลให้กำไรของผู้ผลิตเพิ่มขึ้น ผู้ผลิตก็สามารถเพิ่มการออมได้มากกว่าอัตราการเพิ่มของรายได้ของแรงงาน ดังนั้น จึงทำให้รายจ่ายเพื่อการบริโภคลดลงและการออมเพิ่มขึ้น เมื่อระบบเศรษฐกิจมีการออมเพิ่มและรายจ่ายเพื่อการบริโภคลดลง จึงส่งผลให้รายได้ประชาชาติและการจ้างงานลดหรือการว่างงานเพิ่มขึ้น จึงมีสมมติฐานของทฤษฎีนี้อยู่ที่ว่า ถ้าไม่ต้องการให้มีการว่างงานเพิ่มขึ้นจำเป็นต้องมีอุปสงค์รวมที่เพียงพอ โดยเชื่อว่าการว่างงานจะเกิดขึ้นได้เสมอถ้ารัฐบาลไม่ใช้นโยบายทางเศรษฐกิจ ระบบเศรษฐกิจอาจอยู่ในภาวการณ์จ้างงานไม่เต็มที่ไปเรื่อย ๆ

2.1.2.3 ทฤษฎีการว่างงานของสำนักนีโอคลาสิค (Neoclassical school)

เป็นทฤษฎีที่เชื่อว่าระบบเศรษฐกิจจะมีอัตราการว่างงานตามธรรมชาติอยู่ ซึ่งถูกกำหนดโดยความหนืดตัวของตลาดแรงงาน และการว่างงานเพราะโครงสร้างเศรษฐกิจ การดำเนินนโยบายบริหารอุปสงค์มวลรวมแบบขยายตัว อาจมีผลช่วยลดอัตราการว่างงานลงต่ำกว่าอัตราตามธรรมชาติเฉพาะในระยะสั้นเท่านั้น แต่ในระยะยาวเมื่อแรงงานมีประสบการณ์เกี่ยวกับผลการดำเนินนโยบายการเงินหรือการคลัง แรงงานจะสามารถปรับปรุงพฤติกรรมการเสนอขายแรงงานของตัวเองอย่างถูกต้องทำให้อัตราการว่างงานกลับเข้าสู่อัตราธรรมชาติอย่างเดิม ดังนั้น ตามแนวคิดนี้ การไม่มีงานทำ (Non – employment) จะประกอบไปด้วยเหตุผล ได้แก่ การไม่มีงานทำเพราะผลแห่งการตัดสินใจของครัวเรือนที่ไม่ต้องการให้สมาชิกบางคนทำงานในตลาด การว่างงานในรูปของการลงทุนแสวงหางานทำ และการว่างงานเพราะความบกพร่องในตลาดแรงงาน

นักเศรษฐศาสตร์สำนักนีโอคลาสสิค จึงนำเอาทฤษฎีการแสวงหางานทำมาอธิบายอัตราการว่างงานตามธรรมชาติ โดยอธิบายว่าการที่คนงานต้องแสวงหางานทำและนายจ้างต้องแสวงหา ลูกจ้างที่ดีเนื่องจากงานแต่ละอย่างในตลาดแรงงานมีความแตกต่างกันมาก เช่น งานบางอย่างต้องใช้สติปัญญา บางอย่างต้องใช้ฝีมือ บางอย่างต้องใช้กำลัง เป็นต้น แรงงานก็แตกต่างกันในหลาย ๆ ด้าน แต่แรงงานมักไม่มีข้อมูลเกี่ยวกับงานที่ต้องการทำว่ามีอยู่ที่ไหน เงื่อนไขการจ้างเป็นอย่างไร ดังนั้น เพื่อให้ได้มาซึ่งข้อมูลประกอบการตัดสินใจเลือกคนงานจึงต้องยอมสละเวลา (ยอมว่างงาน) และรายได้เพื่อหางานทำ นอกจากนี้ยังมีเหตุผลอีกประการ กล่าวคือจากการหางานทำและการคัดเลือกแรงงานมักจะมีค่าใช้จ่ายสูง ค่าใช้จ่ายเหล่านี้เป็นต้นทุนจม (Sunk cost) แสดงให้เห็นว่า

Ref. code: 25646304010116LCN

ถ้าแรงงานต้องการยึดงานใด เป็นอาชีพนาน ๆ แรงงานต้องลงทุนแสวงหางานค่อนข้างนาน จนกว่าจะแน่ใจว่าได้งานซึ่งทำให้ค่าใช้จ่ายที่ผ่านมานั้นคุ้มค่า

สำหรับทฤษฎีการแสวงหางานทำ อธิบายการเข้าสู่ตลาดแรงงานในรูปของการมีงานทำ โอกาสที่จะได้ทำงาน การว่างงาน และปัจจัยที่มีความสัมพันธ์กับการมีงานทำและทักษะในการหางาน ได้แก่ การรับข่าวสารแรงงาน โอกาสมีงานทำ โดยมีแนวคิดว่าในสภาพที่ตลาดแรงงานมีการให้ข่าวสารข้อมูลน้อย ทั้งนายจ้างและแรงงานขาดความรู้เกี่ยวกับโอกาสของการจ้างงานในตลาดแรงงาน ผู้ที่หางานทำมักจะไม่เลือกงานแรกที่ได้รับการเสนอเนื่องจากขาดข้อมูลเปรียบเทียบ จึงมักใช้เวลาหางานอื่น และข้อมูลอื่นทำให้การว่างงานเกิดขึ้นเพราะปัญหาการขาดประสิทธิภาพของตลาดแรงงานในการให้ข่าวสาร ดังนั้น สามารถสรุปได้ว่า การว่างงานไม่ได้เกิดจากปัญหาความไม่สมัครใจเพียงอย่างเดียวและอาจเป็นไปได้ที่การว่างงานเป็นจำนวนมากเป็นการว่างงานชั่วคราว และเป็นไปโดยสมัครใจ เพราะใช้เวลาว่างหางานทำถือเป็นการลงทุนเพื่อผลในอนาคต และปัจจัยที่สำคัญที่จะมีผลต่อการตัดสินใจเข้าสู่ตลาดแรงงานก็คือ ระบบข่าวสารแรงงานที่ได้รับ อัตราค่าจ้างที่พอใจ ทักษะในการทำงานและฐานะทางเศรษฐกิจ

2.1.2.4 ทฤษฎีแรงจูงใจ (motivation theory)

Maslow (1943) ได้ตั้งสมมติฐานสิ่งจูงใจจากความต้องการของมนุษย์ไว้ว่า มนุษย์มีความต้องการอยู่เสมอ ถ้าความต้องการนั้นได้รับการตอบสนองแล้วจะไม่เป็นสิ่งจูงใจอีกต่อไป แต่ความต้องการที่ยังไม่ได้รับการตอบสนองนั้นจะเป็นสิ่งจูงใจแทน และแบ่งลำดับขั้นความต้องการของมนุษย์จากระดับต่ำถึงระดับสูง ออกเป็น 5 ขั้น ดังนี้

(1) ความต้องการทางด้านร่างกาย (physiological needs) เป็นความต้องการพื้นฐานเพื่อความอยู่รอดของชีวิต เช่น ความต้องการอาหาร อากาศ ที่อยู่อาศัย ยารักษาโรค และความต้องการทางเพศ เป็นต้น

(2) ความต้องการทางด้านความปลอดภัย (safety needs) ได้แก่ ความต้องการความปลอดภัย ทางด้านร่างกาย เช่น ความปลอดภัยจากอุบัติเหตุ อันตรายต่าง ๆ และความมั่นคงในอาชีพ

(3) ความต้องการที่จะเป็นส่วนหนึ่งของสังคม (belonging needs) ได้แก่ ความต้องการที่จะเข้าร่วม และได้รับการยอมรับในสังคม ความเป็นมิตร และความรักจากเพื่อนร่วมงาน

(4) ความต้องการเห็นคุณค่าของตนเอง (esteem needs) ได้แก่ ความต้องการอยากเด่นในสังคม เป็นที่ยอมรับ เป็นที่ยกย่องสรรเสริญของบุคคลอื่น

Ref. code: 25646304010116LCN

(5) ความต้องการที่จะได้รับความสำเร็จตามความนึกคิดของตนเอง (self - actualization) เป็นความต้องการขั้นสูงสุดของมนุษย์ ที่คนส่วนมากอยากจะเป็นอยากจะได้

2.1.3 ทฤษฎีเกี่ยวกับพฤติกรรมสารสนเทศ

จากการรวบรวมข้อมูลของผู้ศึกษาเกี่ยวกับทฤษฎีพฤติกรรมของประชาชนที่ใช้ระบบสารสนเทศ พบว่า รุจเรขา วิทยาวุฒิกุล (2554) ได้ให้ความหมายของ พฤติกรรมสารสนเทศ (Information behavior) ว่า “เป็นการศึกษาเพื่อทำความเข้าใจวิธีการต่างๆ ที่มนุษย์มีปฏิสัมพันธ์กับสารสนเทศ โดยเฉพาะการสร้าง การแสวงหา และการใช้สารสนเทศ” อีกทั้งยังมีทฤษฎีทางจิตวิทยาเข้ามาเกี่ยวข้อง เนื่องจากความต้องการสารสนเทศเกิดจากแรงกระตุ้นทางกาย ความอยากรู้อยากเห็น และแรงกระตุ้นทางสังคม เช่น ต้องการถูกยอมรับ หรือแรงกระตุ้นที่เกิดจากความต้องการที่จะรู้หรือเข้าใจสิ่งแวดล้อมรอบตัว ซึ่งเป็นความต้องการที่ซับซ้อน

Wilson (2000) ได้ให้ความหมายของพฤติกรรมสารสนเทศ หมายถึง พฤติกรรมของบุคคลหนึ่งที่เชื่อมโยงบุคคลนั้นใหเข้าถึงแหล่งสารสนเทศต่าง ๆ โดยใชชองทางในการเผยแพรจากสื่อต่าง ๆ เช่น โทรทัศน์ อินเตอร์เน็ต ไม่ว่าจะเจตนาหรือไม่ได้เจตนาที่จะได้รับสารสนเทศนั้น ๆ โดยพฤติกรรมนั้นเป็น พฤติกรรมที่ครอบคลุมกิจกรรมสำคัญ 2 กิจกรรม คือ การค้นหาสารสนเทศที่ต้องการด้วยวิธีใด ๆ และ การใช้สารสนเทศหรือการส่งต่อสารสนเทศนั้น ๆ ให้กับผู้อื่น นอกจากนี้ทฤษฎีพฤติกรรมสารสนเทศจะมีความสัมพันธ์ระหว่างพฤติกรรมสารสนเทศกับพฤติกรรมที่เกี่ยวข่องได้แก่ พฤติกรรมการแสวงหาสารสนเทศ (Information-seeking Behavior) หมายถึง การแสวงหาสารสนเทศอย่างมีวัตถุประสงค์ โดยเป็นผลมาจากความต้องการของบุคคลนั้น ซึ่งในระหว่างการแสวงหานั้น บุคคลนั้นจะต้องมีการปฏิสัมพันธ์กับระบบสารสนเทศ ไม่ว่าจะเป็นระบบสารสนเทศโดยมนุษย์ หรือระบบสารสนเทศด้วยคอมพิวเตอร์ และยังมีพฤติกรรมการค้นหาสารสนเทศ (Information Search Behavior) หมายถึง พฤติกรรมที่บุคคลได้ปฏิสัมพันธ์กับระบบสารสนเทศ ทั้งในระดับปฏิบัติ เช่น การใช่เมาส์ หรือในระดับการใช้ความคิด สติปัญญา และความรู้ เช่น การตัดสินใจเลือก หรือพิจารณาว่าสารสนเทศที่ตนเองค้นหานั้นตรงกับความต้องการของตนเองหรือไม่ (รุจเรขา วิทยาวุฒิกุล, 2554)

2.1.4 ทฤษฎีการพยากรณ์และการสร้างแบบจำลองด้วย Machine Learning

ในปัจจุบันมีวิธีการสร้างแบบจำลองเพื่อการพยารณ์ต่าง ๆ ที่หลากหลาย ทั้งการใช้สถิติในรูปแบบดังเดิม การทำเหมืองข้อมูล (Data Mining) หรือการใช้ปัญญาประดิษฐ์ (Artificial Intelligence: AI) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยในงานศึกษานี้จะเลือกใช้วิธีการสร้างแบบจำลองด้วยวิธีการของปัญญาประดิษฐ์

Ref. code: 25646304010116LCN

ปัญญาประดิษฐ์ (Artificial Intelligence) หรือ AI หมายถึง ศาสตร์แขนงหนึ่งของวิทยาศาสตร์คอมพิวเตอร์ที่เกี่ยวข้องกับวิธีการทำให้คอมพิวเตอร์มีความสามารถคล้ายมนุษย์หรือเลียนแบบพฤติกรรมมนุษย์ คือ โปรแกรมซอฟแวร์ (Software) ต่าง ๆ ที่ใช้กับคอมพิวเตอร์โดยเฉพาะความสามารถในการคิดเองได้ โดยผ่านกระบวนการเรียนรู้ต่าง ๆ ที่เรียกว่า Machine Learning (บุญเสริม กิจศิริกุล, 2548) ซึ่ง Machine Learning เป็นส่วนการเรียนรู้ของเครื่องที่ถูกใช้งานเสมือนเป็นสมองของ AI อาจกล่าวได้ว่า AI ใช้ Machine Learning ในการสร้างความฉลาด โดยมักจะใช้เรียกแบบจำลองที่เกิดจากการเรียนรู้ของปัญญาประดิษฐ์ ไม่ได้เกิดจากการเขียนโดยใช้มนุษย์แต่มนุษย์มีหน้าที่เขียนโปรแกรมให้ AI เรียนรู้จากข้อมูลเท่านั้น ที่เหลือเครื่องจัดการเอง Machine Learning เรียนรู้จากสิ่งที่เราส่งเข้าไปกระตุ้น แล้วจดจำเอาไว้เป็นมันสมอง ส่งผลลัพธ์ออกมาเป็นตัวเลขที่ส่งต่อไปแสดงผล หรือให้ AI นำไปแสดงการกระทำ Machine Learning เองสามารถเอาไปใช้งานได้หลายรูปแบบ ต้องอาศัยกลไกที่เป็นโปรแกรม หรือเรียกว่า อัลกอริทึม (Algorithm) ที่มีหลากหลายแบบ โดยมีนักวิทยาศาสตร์ข้อมูล (Data Scientist) เป็นผู้ออกแบบ หนึ่งในอัลกอริทึมที่ได้รับความนิยมสูง คือ Deep Learning ซึ่งถูกออกแบบมาให้ใช้งานได้ง่าย และประยุกต์ใช้ได้หลายลักษณะงาน อย่างไรก็ตาม ในการทำงานจริง Data Scientist จำเป็นต้องออกแบบตัวแปรต่าง ๆ ทั้งในตัวของ Deep Learning เอง และต้องหาอัลกอริทึมอื่น ๆ มาเป็นคู่เปรียบเทียบ เพื่อหาอัลกอริทึมที่เหมาะสมที่สุดในการใช้งานจริง (มัทนา วิบุญศักดิ์, 2021)

ประเภทการเรียนรู้ของ Machine Learning สามารถแบ่งได้เป็น 3 รูปแบบ ดังนี้

(1) Supervised learning หรือการเรียนรู้แบบมีผู้สอน เป็นเทคนิคการเรียนรู้โดยจะแบ่งข้อมูลออกเป็น 2 ชุด ประกอบด้วย ชุดข้อมูลสำหรับสอนให้เครื่องเรียนรู้ (Training Data) เพื่อให้เครื่องทำการเรียนรู้ระหว่างลักษณะข้อมูลที่มี (Feature) กับผลที่ต้องการ และทำการสร้างฟังก์ชันหรือตัวแบบที่ใช้ทำนายหรือบอกผลที่ต้องการ (ฤทธิไกร ทวีเจริญ, 2559) ซึ่งผลที่ต้องการอาจอยู่ในรูปของการถดถอย (Regression) หรือ เป็นการแบ่งแยกประเภท (Classification) และ ชุดข้อมูลที่ใช้ทดสอบ (Test Data) เพื่อใช้วัดประสิทธิภาพและความถูกต้องของฟังก์ชันหรือตัวแบบที่ทำการสร้างขึ้นมาจากชุดข้อมูลที่ใช้ในการเรียนรู้หรือสร้างแบบจำลอง (Training Data) ว่ามีความเหมาะสมในการนำไปใช้กับข้อมูลใหม่ ๆ ที่ยังไม่เคยเห็นได้หรือไม่ ซึ่งจะต้องมีการปรับให้ได้ค่าที่เหมาะสมก่อนนำไปใช้จริงซึ่งจะประกอบด้วยอัลกอริทึม ได้แก่ Linear Regression, Logistic Regression, Decision Tree, Naïve Bayes, Support Vector Machine และ Random Forest เป็นต้น

(2) Unsupervised Learning หรือการเรียรู้แบบไม่มีผู้สอน เป็นเทคนิคการใช้ชุดข้อมูลที่มีทั้งหมดในการทำให้เครื่องทำการเรียนรู้ โดยไม่จำเป็นต้องระบุผลที่ต้องการ หรือ ชนิด

Ref. code: 25646304010116LCN

ประเภทข้อมูลไว้ก่อนที่จะให้เครื่องทำการเรียนรู้ เครื่องจะทำหน้าที่แยกผลที่ต้องการ หรือ ประเภทตลอดจนรูปแบบให้เองจากชุดข้อมูลที่ใช้ในการเรียนรู้ ซึ่งจะประกอบด้วยอัลกอริทึม ได้แก่ K-means Clustering, Gaussian Mixture Model, Hierarchical Clustering, Recommender System และ PCA เป็นต้น

(3) Reinforcement Learning หรือ การเรียนรู้แบบเสริมกำลัง เป็นเทคนิคการเรียนรู้ที่เครื่องจะสนใจต่อสิ่งแวดล้อมที่เกิดจากปฏิสัมพันธ์แบบลองผิดลองถูกระหว่างผู้กระทำบางอย่างกับสิ่งแวดล้อมที่ผู้กระทำมีปฏิสัมพันธ์ด้วย โดยผู้ที่กระทำจะสามารถรับรู้สถานการณ์ของสิ่งแวดล้อมผ่านแต่ละสถานการณ์ และเลือกการกระทำที่ส่งผลต่อสิ่งแวดล้อมนั้น โดยหวังว่าจะได้ผลลัพธ์ที่ดีที่สุด รวมทั้งเรียนรู้ผ่านข้อผิดพลาดในอดีตที่เกิดขึ้น

ในงานศึกษานี้ จะใช้อัลกอริทึมของ Machine Learning ประเภท Supervised learning ในรูปของการถดถอย (Regression) เนื่องจากประเภทของข้อมูลที่ใช้ในการศึกษามีลักษณะเป็นการพยากรณ์เชิงปริมาณ (Quantitative Method) และเลือกใช้อัลกอริทึม 5 วิธี ได้แก่ Linear Regression Model, Optimizable Tree Regression Model, Optimizable Support Vector Regression Model, Neural Network Model และ Optimizable Ensemble Model ซึ่งเป็นวิธีที่ได้รับความนิยม และการยอมรับว่ามีความแม่นยำระดับสูง เหมาะกับการวิเคราะห์ตัวแปรทางเศรษฐกิจ โดยแต่ละวิธีต่างมีลักษณะการใช้งาน ข้อดีและข้อเสีย ตลอดจนคุณภาพความแม่นยำที่ได้แตกต่างกันไป จึงนำมาศึกษาและเปรียบเทียบผลจากการพยากรณ์อัตราการว่างงาน เพื่อหาอัลกอริทึมที่มีประสิทธิภาพในการพยากรณ์มากที่สุด สามารถอธิบายรายละเอียดได้ ดังนี้

2.1.4.1 แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model)

Linear Regression หรือ การวิเคราะห์การถดถอยเชิงเส้น เป็นอัลกอริทึมรูปแบบหนึ่งของ Machine Learning ประเภท Supervised Learning หรือ การเรียนรู้แบบมีผู้สอนชนิดแบบ Statistical Regression ที่ต้องใส่ชุดข้อมูลเข้าไปให้โปรแกรมเรียนรู้ก่อน โดยโปรแกรมจะนำตัวแปรต้นและตัวแปรตามไปคำนวณด้วยสถิติทางคณิตศาสตร์ แล้วก็จะได้ข้อมูลกลับมาเป็นตัวเลข

อัลกอริทึม Linear Regression จะแสดงความสัมพันธ์ระหว่างตัวแปรต้น (x) กับตัวแปรตาม (y) และใช้กลไกทางคณิตศาสตร์ มาเรียนรู้ข้อมูล Training Data เพื่อสร้างแบบจำลองของข้อมูลขึ้นมา และใช้ในการพยากรณ์คำตอบหรือ y ที่ต้องการ โดยมีสมการการถดถอยเชิงเส้นอย่างง่าย ดังนี้

Ref. code: 25646304010116LCN

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon$$

โดยที่ y คือ ตัวแปรตาม

x คือ ตัวแปรอิสระ

β_0 และ β_1 คือ สัมประสิทธิ์ของสมการถดถอย

ε คือ ค่าความผิดพลาด

ในการสร้างสมการการถดถอยเชิงเส้นอย่างง่ายที่กล่าวมาข้างต้นจะสร้างจาก Hypothesis function ซึ่งเป็นฟังก์ชันที่เป็นสมมติฐาน ใช้เพื่อหาความสัมพันธ์ของข้อมูล ซึ่งอยู่ในรูปแบบของสมการ

$$\hat{y} = wx + b$$

โดยที่ $\hat{y}$ คือ ค่าที่ต้องการพยากรณ์

w คือ ค่าน้ำหนัก (weight) ของตัวแปร x ซึ่งกำหนดความชันของฟังก์ชัน

x คือ ตัวแปรอิสระ

b คือ จุดตัดแกน y (หรือ intercept coefficient)

หน้าที่ของ Machine Learning คือการหาตัวแปร w และ b ซึ่งจะใช้ประกอบกันเป็น Hypothesis function ที่มีความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับข้อมูล Training Data สำหรับ Hypothesis function แต่ละฟังก์ชัน จะมี Cost function คู่อยู่เสมอ ซึ่งทำหน้าที่วัดค่า y ที่พยากรณ์ได้จาก Hypothesis function มีความคลาดเคลื่อนกับค่า y ที่มีข้อมูลอยู่จริงเท่าใด (ชิตพงษ์ กิตตินราดร, 2562) ใน Linear regression จะใช้ฟังก์ชันต่อไปนี้เป็น Cost function

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

โดยที่ $J(w,b)$ คือ ผลความต่างระหว่างค่าพยากรณ์กับค่าจริงที่สอดคล้องกับตัวแปร w และ b

m คือ จำนวนข้อมูล Training Data ทั้งหมด

$(\hat{y}_i - y_i)^2$ คือ ผลต่างระหว่างค่า y ที่พยากรณ์ได้ กับค่า y จริงของแต่ละตัวยกกำลัง 2

อย่างไรก็ตาม Cost function สามารถเรียกว่า Mean Squared Error หรือ MSE ซึ่งหลักการของอัลกอริทึมจะต้องการหา Cost function ต่ำที่สุดที่จะเป็นไปได้ โดยจะใช้กระบวนการที่เรียกว่า Gradient descent เพื่อหาตัวแปร w และ b ที่จะทำให้ได้ J ที่ต่ำที่สุด จาก

Ref. code: 25646304010116LCN

การหาอนุพันธ์ (Derivative) ของฟังก์ชัน $\mathbf{J(w)}$ เมื่อเปรียบเทียบกับ $\mathbf{w}$ และ Machine Learning จะทำซ้ำจนกระทั่งผลลัพธ์ล่าสุดไม่เปลี่ยนแปลงจากผลลัพธ์ครั้งก่อน จะได้สมการ

$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)$$

$$w = w - \alpha \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)\, x_i$$

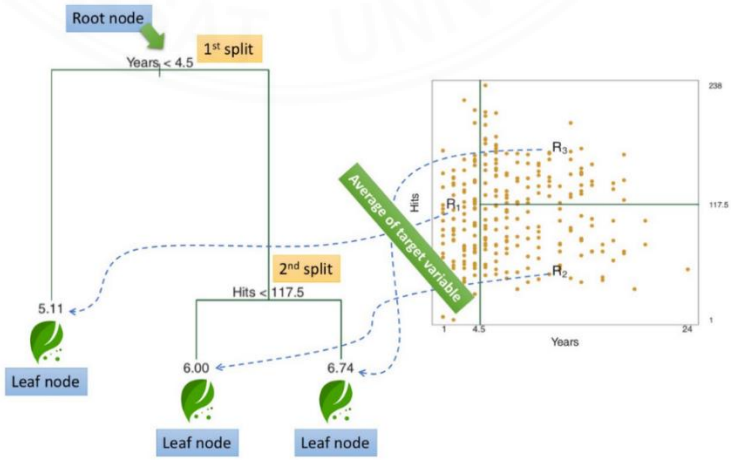
ซึ่งเป้าหมายของ Gradient descent คือการหาค่าตัวแปรของแบบจำลองที่ทำให้มีผลลัพธ์ $\mathbf{y}$ ที่คลาดเคลื่อนกับค่า $\mathbf{y}$ ที่แท้จริงให้น้อยที่สุด กล่าวคือจะได้แบบจำลองที่เป็นตัวแทนของชุดข้อมูลนี้ได้ดีที่สุด

2.1.4.2 แบบจำลองต้นไม้ตัดสินใจ (Optimizable Decision Tree Regression Model)

Decision Tree หรือ ต้นไม้ตัดสินใจ เป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ที่มีลักษณะคล้ายกับต้นไม้ จะมีการสร้างกฎต่าง ๆ ขึ้นเพื่อใช้ในการตัดสินใจ และเป็น Machine Learning ประเภท Supervised Learning หรือ การเรียนรู้แบบมีผู้สอน คือ สามารถสร้างแบบจำลองการจัดหมวดหมู่จากกลุ่มตัวอย่างของข้อมูลที่กำหนด กล่าวคือ Training Data ได้อัตโนมัติ และสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้

ภาพที่ 2.1

การแบ่งข้อมูลของ Decision Tree Regression



หมายเหตุ. The three-region partition for the Hitters data set from the regression tree. โดย James et al, 2021, (https://hastie.su.domains/ISLR2/ISLRv2_website.pdf)

Ref. code: 25646304010116LCN

หลักการของ Decision Tree ในภาพที่ 2.1 จะแบ่งข้อมูลออกทีละ 2 ส่วน (Recursive binary split) จาก node ล่างสุดของต้นไม้เรียกว่า โหนดราก (Root node) และไล่ขึ้นมาเรื่อย ๆ จนถึงโหนดใบ (Leaf node) ตามภาพที่ 2.1 ด้านซ้าย และทำการพยากรณ์ค่าตัวแปรเป้าหมาย (Target variable) ด้วยวิธีการหาค่าเฉลี่ยของโหนดตัวแปรเป้าหมาย โดยการแบ่งข้อมูลจากโหนดรากจนถึงโหนดใบ และทำจนกว่าจะได้เงื่อนไขที่กำหนด เช่น ความลึกของต้นไม้ไม่เกิน 10 ชั้น (Max depth) หรือ จำนวนข้อมูลในแต่ละกลุ่มที่แบ่งออกมาในโหนดใบมีจำนวนขั้นต่ำ 5 ข้อมูล (Min sample)

สำหรับการแบ่งข้อมูลแต่ละโหนด เริ่มจากเลือกข้อมูลมาทำการคัดกรองข้อมูลด้วยค่าของข้อมูลที่เลือกมา จากนั้นหาจุดแบ่งที่เป็นไปได้ทั้งหมดจากจำนวนข้อมูลทั้งหมด ซึ่งในการแบ่งข้อมูลแต่ละแบบที่เป็นไปได้ จะคำนวณหาค่า Residual Sum of Squares (RSS) จากการพยากรณ์ตัวแปรเป้าหมายด้วยค่าเฉลี่ยของตัวแปรเป้าหมายในแต่ละกลุ่ม และทำการแยกข้อมูลที่มีค่า RSS น้อยที่สุดออกมา เมื่อสิ้นสุดการแยกแล้ว จะพยากรณ์ค่าที่ต้องการจากค่าเฉลี่ยของค่าพยากรณ์ภายในโหนดของตัวเอง สำหรับโหนดที่ยังสามารถแยกได้ต่อ จะเรียกค่าสุดท้ายของการแยกที่ให้ค่าพยากรณ์มาเป็นค่าเฉลี่ยของค่าพยากรณ์นั้นในแต่ละกลุ่มว่า โหนดใบ ทั้งนี้ RSS มีสูตรคำนวณได้ ดังนี้

$$\mathrm{RSS} = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \widehat{y_{R_j}})^2$$

โดยที่ R_j คือ แต่ละกลุ่มของข้อมูลที่ถูกแบ่งออกมา เป็นทั้งหมด J กลุ่ม

y_i คือ ค่าที่ต้องการพยากรณ์

$\widehat{y_{R_j}}$ คือ ค่าพยากรณ์ในแต่ละกลุ่มที่คำนวณมาจากค่าเฉลี่ยของค่าพยากรณ์ในกลุ่มนั้น ๆ

2.1.4.3 แบบจำลองซับพอรตเวกเตอร์รีเกรสชัน (Optimizable Support Vector Regression Model: SVR)

Support Vector Regression เป็นส่วนหนึ่งของอัลกอริทึม Support Vector Machine ซึ่งเป็น Machine Learning ประเภท Supervised Learning หรือ การเรียนรู้แบบมีผู้สอน โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด (จารุวิทย์ ประทานชีวินทร์, 2562) Support Vector Regression ทำงานบนหลักการของ Support Vector Machine โดยมีความแตกต่างเล็กน้อย กล่าวคือการกำหนดจุดข้อมูล จะพยายามหาเส้นโค้งแต่เนื่องจากเป็นอัลกอริทึมการถดถอยแทนที่จะใช้เส้นโค้งเป็นขอบเขตการตัดสินใจ จะใช้เส้นโค้งเพื่อ

Ref. code: 25646304010116LCN

ค้นหาการจับคู่ระหว่างเวกเตอร์กับตำแหน่งของเส้นโค้ง Support Vectors ช่วยในการกำหนดการจับคู่ที่ใกล้เคียงที่สุดระหว่างจุดข้อมูลและฟังก์ชันที่ใช้ (Pedamkar, 2019)

อัลกอริทึมของ Support Vector Regression ประกอบด้วยส่วนต่าง ๆ ดังนี้ และแสดงในภาพที่ 2.2

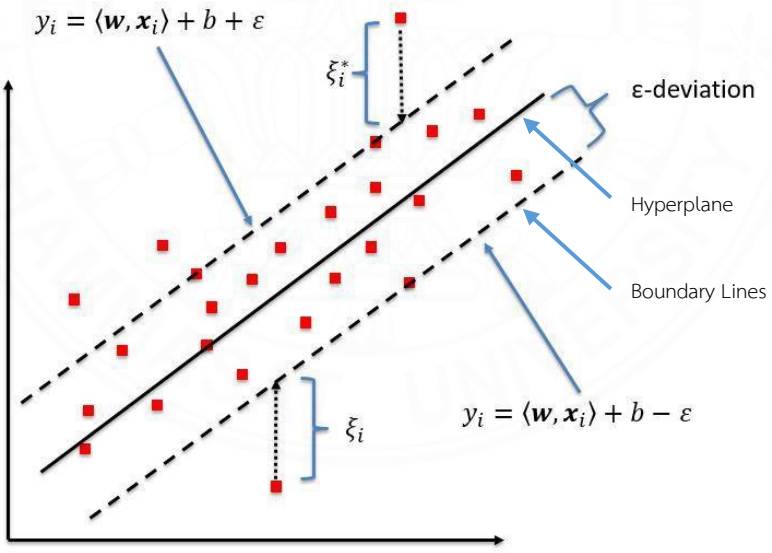
(1) Hyperplane: เป็นเส้นแบ่งระหว่างสองข้อมูลในมิติที่สูงกว่ามิติจริง ใน SVR ถูกกำหนดให้เป็นเส้นที่ช่วยในการพยากรณ์ค่าเป้าหมาย

(2) Kernel: ใน SVR สามารถสร้างเส้นแบ่งขอบเขตการตัดสินใจในรูปแบบที่ไม่ใช่แส้นตรงได้ ฟังก์ชันนี้เรียกว่า Kernel ประเภทของ Kernel ที่ใช้ใน SVR คือ Sigmoidal Kernel, Polynomial Kernel และ Gaussian Kernel เป็นต้น

(3) Boundary Lines: คือเส้นสองเส้นที่ลากรอบ Hyperplane ที่ระยะห่าง Epsilon (ζ) ใช้เพื่อสร้างระยะขอบระหว่างจุดข้อมูล

ภาพที่ 2.2

การแบ่งขอบเขตของ Support Vector Regression



หมายเหตุ. โดย Educba, 2018, (https://www.educba.com/support-vector-regression/)

อัลกอริทึม Support Vector Regression มีหลักการคือต้องการหาความสัมพันธ์เชิงเส้นระหว่างอินพุตเวกเตอร์ (Input vector) ในมิติขนาด $\mathrm{n}(\mathrm{X} \in \mathrm{R}^{\mathrm{n}})$ และตัวแปรเอาท์พุต (Output variables) ในมิติ $(\mathrm{y} \in \mathrm{R})$ ซึ่งแสดงในรูปสมการถดถอยได้

Ref. code: 25646304010116LCN

$$f(x) = w^T x + b$$

โดยที่ w คือ น้ำหนัก (Weight) ของเวกเตอร์

b คือ bias

ซึ่ง w และ b หาได้ด้วย Hypothesis function ของ Support Vector Regression จะทำการหาค่าคาดเคลื่อนที่ต่ำที่สุด (Minimize error) ด้วย Epsilon insensitive ที่นำเสนอโดย Vapnik (1998) เพราะเป็นฟังก์ชันสูญเสีย (Loss function) ที่นิยมใช้ในงานวิจัยอย่างแพร่หลาย จะแสดงในรูปแบบของ Absolute term หลักการคือการสร้างพื้นที่ขึ้นมา เพื่อพิจารณาข้อมูลที่หลุดออกนอก Boundary Lines โดยจะกำหนดให้เป็น Error ของแบบจำลอง ดังภาพที่ 2.2 และเขียนสมการได้ ดังนี้

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}|\zeta_i + \zeta_i^*|$$

โดยที่ w คือ น้ำหนัก (Weight) ของเวกเตอร์

C คือ เป็นจำนวนเต็มที่มีค่าคงที่เพื่อหาค่าคาดเคลื่อนที่ต่ำที่สุด

m คือ จำนวนข้อมูล Training Data

ζ คือ ระยะห่างระหว่างจุดข้อมูลกับ Boundary Lines

การใช้เทคนิค SVR พยากรณ์ตัวแปรเอาต์พุตจากอินพุตเวกเตอร์ จะมีการสร้างท่อ ζ (Epsilon tube) โดยใช้ฟังก์ชันสูญเสีย (Loss function) แบบ Epsilon insensitive loss function แสดงได้ ดังนี้

$$|y_i - f(x_i)|_\zeta = \begin{cases} 0 & \text{if} |y_i - f(x_i)|_\zeta \le \zeta \\ |y_i - f(x_i)|_\zeta - \zeta & \text{otherwise} \end{cases}$$

หรือสามารถเขียนให้อยู่ในรูปการใช้ตัวคูณลากรองจ์ (Lagrange multipliers) ดังนี้

$$\text{Maximize } L_p(\alpha_i, \alpha_i^*) = -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T x_j$$

Ref. code: 25646304010116LCN

$$-\zeta\sum_{i=1}^{m}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{m}(\alpha_i-\alpha_i^*)y_i$$

$$\text{Subject to}\begin{cases}\sum_{i=1}^{m}(\alpha_i-\alpha_i^*)=0\\ 0\le\ \alpha_i\le C,\quad i=1,\dots,m\\ 0\le\ \alpha_i^*\le C,\quad i=1,\dots,m\end{cases}$$

โดยที่ α_i , α_i^* คือ ตัวคูณ Lagrange

C คือ เป็นจำนวนเต็มที่มีค่าคงที่เพื่อหาค่าคาดเคลื่อนที่ต่ำที่สุด

ζ คือ ความคลาดเคลื่อนของชุดข้อมูล Training Data

m คือ จำนวนของซัพพอร์ตเวกเตอร์ (Support vector)

ทั้งนี้ ถ้า $\alpha_i , \alpha_i^* > 0$ จะเป็นอินพุตเวกเตอร์ที่เป็นซัพพอร์ตเวกเตอร์ด้วย และถ้า $\alpha_i , \alpha_i* = 0$ เป็นอินพุตเวกเตอร์ที่ไม่ใช่ซัพพอร์ตเวกเตอร์ หลังจากคำนวณค่า α_i และ α_i^* จะสร้างสมการ SVR ได้ ดังนี้

$$f(x)=\ w_0^Tx+b=\sum_{i=1}^{m}(\alpha_i-\alpha_i^*)w_i^Tx+b$$

โดยที่ $w_0=\sum_{i=1}^{m}(\alpha_i-\alpha_i^*)w_i$

แต่ในกรณีที่อยู่ในรูปสมการถดถอยแบบไม่เป็นเชิงเส้นสามารถส่งอินพุตเวกเตอร์ไปยังมิติที่สูงขึ้น (High dimensional feature space) โดยใช้เคอร์เนลฟังก์ชัน (Kernel function) ซึ่งเคอร์เนลฟังก์ชันที่นิยมใช้ใน SVR มีดังนี้

(1) ลีเนียร์เคอร์เนล (Linear kernel)

$$k(x_i,x)=\ x_i^Tx$$

(2) โพลีโนเมียลเคอร์เนล (Polynomial kernel)

$$k(x_i,x)=(1+x_i\cdot x_j)^d$$

(3) เกาส์เซียนเคอร์เนล (Gaussian (RBF))

$$k(x_i,x)=\ \exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)$$

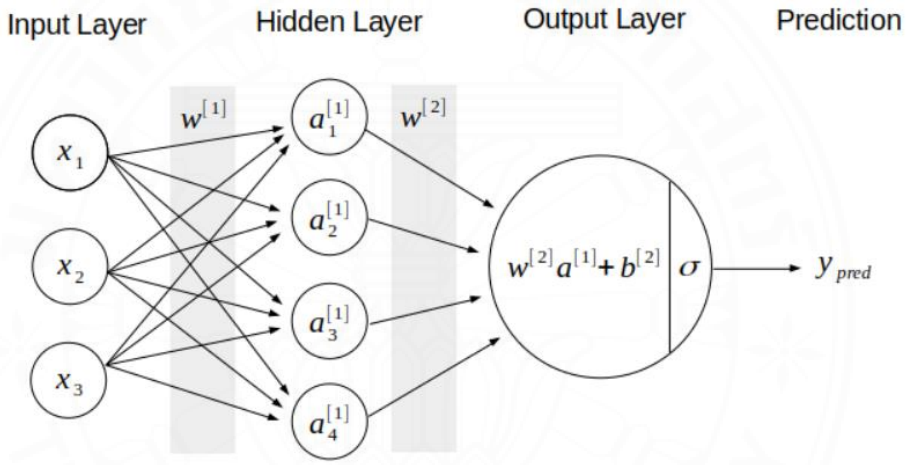
Ref. code: 25646304010116LCN

2.1.4.4 แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN)

Neural Network หรือ โครงข่ายประสาทเทียม เป็น Machine Learning ประเภท Supervised Learning หรือ การเรียนรู้แบบมีผู้สอน โครงสร้างการเรียนรู้ (Learning architecture) ของ Neural network ต่างจากโครงสร้างการเรียนรู้ของ Machine Learning อัลกอริทึมแบบทั่วไป ดังนี้

ภาพที่ 2.3

องค์ประกอบของ Neural Network



หมายเหตุ. โดย ชิตพงษ์ กิตตินราดร, 2563, (https://guopai.github.io/ml-blog14.html)

จะเห็นว่าส่วนประกอบของโครงข่ายประสาทเทียมมี 4 ส่วน ได้แก่ ชั้นข้อมูลเข้า (Input layer), ชั้นซ่อน (Hidden layer), ชั้นข้อมูลออก (Output layer) และ ค่าพยากรณ์ (Prediction) ทั้งนี้ ชั้นซ่อนอาจมีมากกว่าหนึ่งชั้น ในแต่ละชั้นจะประกอบด้วยโหนดหรือเซลล์ประสาทหนึ่งเซลล์หรือมากกว่า และข้อมูลจะถูกส่งจากชั้นข้อมูลเข้าไปถึงชั้นข้อมูลออก โดยไม่มีการส่งย้อนกลับ หรือแม้กระทั่งโหนดในชั้นเดียวกันก็ไม่มีการเชื่อมต่อกัน (กานต์สินี เจริญกิจวัชรชัย, 2561)

ในแต่ละชั้นจะประกอบไปด้วยโหนด (Node) หลายโหนดโยงกันเป็นโครงข่าย ประกอบด้วย

(1) โหนดขาเข้า (Input Nodes) คือ โหนดที่ให้ข้อมูลจากภายนอกแก่โครงข่าย ทำหน้าที่ป้อนข้อมูลให้กับโครงข่ายไปสู่ปมซ่อนเร้นภายในชั้นซ่อนเร้นของโครงข่ายภายใน

Ref. code: 25646304010116LCN

(2) โหนดซ่อนเร้น (Hidden Nodes) คือ โหนดที่คำนวณและประมวลผลข้อมูลที่ได้รับจากโหนดขาเข้า แล้วส่งผลลัพธ์ที่ได้ไปสู่โหนดขาออก ซึ่งโหนดซ่อนเร้นทุก ๆ โหนดก็จะอยู่ในชั้นซ่อนเร้นของโครงข่าย

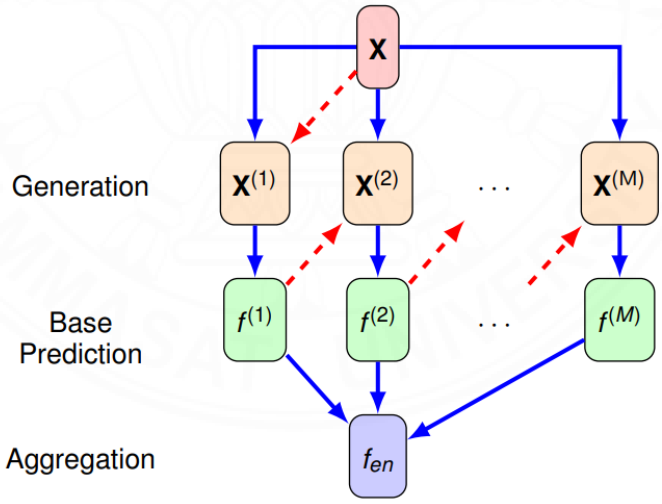
(3) โหนดขาออก (Output Node) คือ โหนดที่แสดงผลการคำนวณภายในโครงข่าย โหนดขาออกจะอยู่ในชั้นขาออกของโครงข่าย

2.1.4.5 แบบจำลองแบบเรียงซ้อน (Optimizable Ensemble Model: EM)

Ensemble หรือ การวิเคราะห์แบบเรียงซ้อน เป็น Machine Learning ประเภท Supervised Learning หรือ การเรียนรู้แบบมีผู้สอน ซึ่ง Ensemble มีทฤษฎีหลักที่อยู่เบื้องหลังคือการจำแนกของ bias - variance - covariance เพื่อปรับปรุงประสิทธิภาพของการพยากรณ์ สิ่งสำคัญในแบบจำลองนี้ คือ ความหลากหลาย ซึ่งรวมถึงความหลากหลายของข้อมูล พารามิเตอร์ และโครงสร้าง เพื่อปรับให้เหมาะสมกับความหลากหลายของวัตถุประสงค์ (Ren et al., 2016)

ภาพที่ 2.4

การแบ่งแบบจำลองของ Ensemble Regression



หมายเหตุ. โดย Ren et al., 2016, (https://ieeexplore.ieee.org/document/7379058)

หลักการทำงานของ Ensemble แสดงได้ในภาพที่ 2.4 โดยที่ $\mathbf{x}$ คือ ชุดข้อมูลที่ใช้ป้อนเข้าในแบบจำลอง $\mathbf{x}^{(i)}, i \in \{1,...,M\}$ คือ ชุดข้อมูลที่สร้างขึ้นเพื่อเป็นตัวจำแนก $\mathbf{f}^{(i)}$ คือ ค่าการพยากรณ์ในแต่ละแบบจำลอง $\mathbf{f}_{en}$ คือ นำค่า $\mathbf{f}^{(i)}$ มารวมกันเป็นฟังก์ชันเพื่อช่วยตัดสินใจให้

Ref. code: 25646304010116LCN

ผลลัพธ์ออกมาดีที่สุด ซึ่งแบบจำลองที่นำมาใช้ สามารถใช้อัลกอริทึมของ Machine Learning หลายรูปแบบ ได้แก่ Linear Regression, Logistic Regression, Naïve Bay, Decision Tree และอื่น ๆ

2.2 วรรณกรรมปริทัศน์

จากการทบทวนวรรณกรรมปริทัศน์ของงานวิจัยที่เกี่ยวข้องในอดีต พบว่า การศึกษาเกี่ยวกับความสัมพันธ์และการประมาณการอัตราการว่างงานในประเทศไทยที่มีการนำ Google Trends มาประยุกต์ใช้ร่วมด้วยมีอยู่เป็นจำนวนน้อย ผู้ศึกษาจึงทำเลือกการทบทวนวรรณกรรมแบ่งออกเป็น 3 ส่วน ประกอบด้วย

(1) งานวิจัยที่เกี่ยวข้องที่มีการนำ Google Trends มาประยุกต์ใช้ในการหาความสัมพันธ์ และ/หรือ ประมาณการตัวชี้วัดทางเศรษฐกิจ

(2) งานวิจัยที่เกี่ยวข้องที่มีการนำ Google Trends มาประยุกต์ใช้ในการหาความสัมพันธ์ และ/หรือ ประมาณการตัวแปรอื่น ๆ

(3) งานวิจัยที่เกี่ยวข้องที่มีการนำแบบจำลองด้านปัญญาประดิษฐ์มาประยุกต์ใช้ประมาณการตัวแปรต่าง ๆ

นอกจากนี้ ยังได้มีการรวบรวมงานวิจัยจากต่างประเทศด้วย โดยมีรายละเอียดดังนี้

Ref. code: 25646304010116LCN

2.2.1 งานวิจัยที่เกี่ยวข้องที่มีการนำ Google Trends มาประยุกต์ใช้กับตัวชี้วัดทางเศรษฐกิจ

ตารางที่ 2.1

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
ธนัตถ์พราว ลีลารัตน์ (2562)	การเปรียบเทียบความสัมพันธ์ระหว่างอัตราการว่างงานไทยกับข้อมูลที่เกี่ยวข้องบนอินเทอร์เน็ตด้วยกูเกิลเทร์นดส์ (Google Trends) และทวิตเตอร์ (Twitter)	วัตถุประสงค์: เพื่อศึกษาความสัมพันธ์ระหว่างอัตราการว่างงานจากข้อมูลใน Google Trends และ Twitter ของประเทศไทยในแต่ละเดือน และนำข้อมูลที่ได้จากกูเกิลเทร์นดส์และทวิตเตอร์มาใช้ประโยชน์ในอนาคต วิธีการศึกษา: เก็บข้อมูลรายเดือนเป็นจำนวน 72 เดือน จาก Google Trends และ Twitter ซึ่งประกอบด้วยข้อมูล 6 คำ ได้แก่ หางาน สมัครงาน ว่างงาน สัมภาษณ์งาน ประกันสังคม เงินทดแทน และเก็บข้อมูลอัตราการว่างงานรายเดือนของไทย โดยมีการกำหนดตัวแปรอิสระ คือ Google Trends ที่แสดงถึงข้อมูลระดับความสนใจในการค้นหาคำที่เกี่ยวข้องกับอัตราการว่างงาน และ Twitter ที่แสดงค่าร้อยละของจำนวนทวีตคำที่มีการเกี่ยวข้องกับอัตราการว่างงาน ขณะที่ตัวแปรตาม คือ อัตราการว่างงานไทยรายเดือน จากนั้นนำข้อมูลที่ได้ไปหาความสัมพันธ์แบบเพียร์สัน (Pearson Correlation Coefficient) ซึ่งจะใช้ชุดข้อมูลทั้งในช่วงเวลาเดียวกัน และในช่วงเวลาที่ต่างกัน (Lag) เพื่อเปรียบเทียบความสัมพันธ์ระหว่างข้อมูลจากกูเกิลเทร์นดส์และทวิตเตอร์ จากนั้นนำค่าที่มีสัมประสิทธิ์สหสัมพันธ์สูงสุดมาใช้ในแบบจำลอง โดยมีแบบจำลอง 3 รูปแบบ ประกอบด้วย (1) แบบจำลองที่มีแค่ตัวแปรอัตราการว่างงาน ซึ่งจะใช้ตัวแบบ ARIMA (Autoregressive Integrated Moving Average) (2) แบบจำลองที่มีตัวแปร Google Trends ที่มี

Ref. code: 25646304010116LCN

ตารางที่ 2.1

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
		ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันสูงสุดกับอัตราการว่างงานไทย (3) แบบจำลองที่มีตัวแปร Twitter ที่มีค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันสูงสุดกับอัตราการว่างงานไทย โดยแบบจำลองที่ (2) และ (3) จะใช้ตัวแบบ ARIMAX (Autoregressive Integrated Moving Average With Explanatory Variable) ผลการศึกษา: ข้อมูลอัตราการว่างงานไทยมีความสัมพันธ์กับข้อมูลที่ได้จาก Google Trends และ Twitter ไปในทิศทางเดียวกัน แต่มีข้อมูลจาก Twitter จำนวน 3 คำ ที่มีความสัมพันธ์ในทิศทางตรงกันข้าม คือ สมัครงาน ว่างงาน และประกันสังคม ส่วนคำที่มีระดับความสัมพันธ์กับข้อมูล Google Trends มากที่สุดจะเป็นคำว่า ว่างงาน ที่มีค่าสัมประสิทธิ์สหสัมพันธ์ที่ 0.468 และ คำที่มีระดับความสัมพันธ์กับข้อมูล Twitter มากที่สุดจะเป็นคำว่า สัมภาษณ์งาน ที่มีค่าสัมประสิทธิ์สหสัมพันธ์ที่ 0.2349 และมีการศึกษาเกี่ยวกับช่วงเวลาที่ต่างกัน พบว่า ความสัมพันธ์ที่ lag ที่ 1 ของข้อมูล Google Trends และ Twitter มีความสัมพันธ์กับอัตราการว่างงานไทยมากกว่าที่ lag ที่ 0 และ lag ที่ -1 เล็กน้อย ซึ่งแสดงให้เห็นว่า ข้อมูลจาก Google Trends และ Twitter มีลักษณะเป็นตัวชี้วัดตาม (Lagging Indicator) แต่อย่างไรก็ตามเมื่อใช้ข้อมูลที่มีความนิ่งและไม่ผันแปรตามฤดูกาล โดยใช้ lag ที่ 1 และ lag ที่ 12 ร่วมกัน พบว่า อัตราการว่างงานไทยมีความสัมพันธ์กับข้อมูลจาก Google Trends คำว่า เงินทดแทน มากที่สุด และมีความสัมพันธ์กับข้อมูลจาก Twitter คำว่า สมัครงาน มากที่สุด ซึ่งขัดแย้งกับข้อมูลจริง จึงเป็นไปได้ว่า ข้อมูลจริงนั้นอาจมี

Ref. code: 25646304010116LCN

ตารางที่ 2.1

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
		ความสัมพันธ์กับเวลา หรือมีความสัมพันธ์ปลอม (Spurious Relationship) อยู่ ดังนั้น ในการศึกษานี้ จึงนำข้อมูลที่ผ่านการแปลงข้อมูลให้มีความนิ่งและไม่มีความผันแปรจากฤดูกาลแล้ว และมีค่าความสัมพันธ์สูงสุดมาใช้ในการพยากรณ์ ส่วนการนำข้อมูลมาใช้ในการพยากรณ์อัตราการว่างงานไทยโดยเลือกข้อมูลตัวแปรต้นจากข้อมูลที่มีความสัมพันธ์ระหว่างกันสูงสุดจากข้อมูลที่มีความนิ่งและไม่ผันแปรตามฤดูกาล โดยใช้ตัวแบบ ARIMA และ ARIMAX พบว่า ไม่มีความเหมาะสม เพราะไม่ได้ส่งผลให้การพยากรณ์มีความคลาดเคลื่อนลดน้อยลง แต่อาจส่งผลให้ค่าสหสัมพันธ์ที่ได้ระหว่างค่าจริงและค่าพยากรณ์เพิ่มขึ้น
ประภัสสร แสวงสุขสันต์ (2561)	การติดตามภาวะเศรษฐกิจด้วยข้อมูลปริมาณการ Search จาก Google	วัตถุประสงค์: เพื่อใช้ข้อมูลจากอินเทอร์เน็ตในการพัฒนาตัวชี้วัดทางเศรษฐกิจต่าง ๆ วิธีการศึกษา: เก็บข้อมูลจาก Google Correlate และ Google Trends ที่เป็นข้อมูลรายเดือน โดยจะเลือกคำที่มีค่า Correlation ในระดับสูงกับข้อมูลที่สนใจ และใช้เครื่องมือ Principle Component Analysis (PCA) ในการจัดกลุ่มความสัมพันธ์ของข้อมูล ซึ่งจะใช้ศึกษาตัวชี้วัดทางเศรษฐกิจต่าง ๆ ได้แก่ เครื่องชี้วัดภาวะรายได้ของครัวเรือน เครื่องชี้วัดความเชื่อมั่นของครัวเรือน เครื่องชี้วัดการบริโภคภาคเอกชนในหมวดสินค้าคงทน และเครื่องชี้วัดจำนวนผู้ว่างงาน ผลการศึกษา: สามารถแบ่งตามตัวชี้วัดทางเศรษฐกิจได้ดังนี้ (1) ตัวชี้วัดภาระรายได้ของครัวเรือน คำว่า LTM + RMF มีค่า Correlation สูงสุดที่ 0.74 (2) ตัวชี้วัดความเชื่อมั่นของครัวเรือน คำว่า ข่าวหุ้น มีค่า Correlation สูงสุดที่ 0.85 (3) ตัวชี้วัดการบริโภคภาคเอกชนในหมวดสินค้าคงทน

Ref. code: 25646304010116LCN

ตารางที่ 2.1

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
		คำว่า ทะเบียนรถ มีค่า Correlation สูงสุดที่ 0.66 (4) ตัวชี้วัดจำนวนผู้ว่างงาน คำว่า หางาน มีค่า Correlation สูงสุดที่ 0.77
Nakavachara and Lekfuangfu (2017)	Predicting the Present Revisited: The Case of Thailand	วัตถุประสงค์: เพื่อนำ Google Trends ไปใช้ในการคาดการณ์เครื่องมือชี้วัดทางเศรษฐกิจต่าง ๆ แบบ Nowcasting วิธีการศึกษา: มีแบบจำลองประกอบด้วย (1) ข้อมูลจากการลงทะเบียนการว่างงาน (Unemployment Registration) ที่เป็นการว่างงานจากการถูกเลิกจ้างหรือไล่ออก และ (2) อัตราการว่างงาน จากนั้นนำข้อมูลที่ได้มาหาค่าความสัมพันธ์กับคำค้นหาใน Google Trends โดยจะเลือกคำค้นหาใน Google ที่มีค่าสหสัมพันธ์อยู่ในกลุ่มที่สูงมาเป็นตัวแปรในแบบจำลอง Seasonal Autoregressive Integrated Moving Average (SARIMA) เพื่อคาดการณ์แบบ Nowcasting ผลการศึกษา: ส่วนของแบบจำลองข้อมูลจากการลงทะเบียนการว่างงานจะใช้คำว่า ตกงาน ซึ่งเป็นคำที่มีค่าสหสัมพันธ์สูงสุด นำมาใช้ร่วมกับตัวแปรอื่น ซึ่งสามารถอธิบายการลงทะเบียนการว่างงานได้มากขึ้น ขณะที่แบบจำลองอัตราการว่างงานจะใช้คำว่า สมัครงาน ที่มีค่าสหสัมพันธ์สูงสุด จะทำให้สามารถอธิบายแบบจำลองได้มากขึ้น แต่ยังคงมีปัญหาของฤดูกาล (Seasonal) และความไม่นิ่งของข้อมูล (Non-stationary) ถึงแม้ว่าจะใช้ Lag 12 แต่ก็ยังไม่สามารถอธิบายแบบจำลองได้ดีขึ้น ข้อมูลอัตราการว่างงานจึงไม่เหมาะสมแบบจำลองนี้

Ref. code: 25646304010116LCN

ตารางที่ 2.1

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
McLaren and Shanbhogue (2011)	Using Internet Search Data as Economic Indicator	วัตถุประสงค์: เพื่อใช้คำค้นหาจาก Internet Search Data เป็นหนึ่งในตัวแปรในแบบจำลองของการพยากรณ์ตลาดแรงงานและตลาดที่อยู่อาศัย วิธีการศึกษา: เก็บข้อมูลในตลาดแรงงานและตลาดที่อยู่อาศัยในประเทศอังกฤษเพื่อสร้างแบบจำลองในการพยากรณ์ของแต่ละตลาด และใช้ตัวแปรต้นเป็นคำค้นหาจาก Internet Search Data และใช้แบบจำลอง Autoregressive Regression (AR) ในการอธิบาย ผลการศึกษา: สร้างแบบจำลองพยากรณ์อัตราการว่างงานโดยใช้คำค้นหาจาก Internet Search Data คือคำว่า JSA (Jobseeker's allowance) เป็นหนึ่งในตัวแปรที่ใช้ในการพยากรณ์ซึ่งสามารถอธิบายแบบจำลองได้ดีขึ้น ส่วนแบบจำลองเพื่อพยากรณ์ราคาในตลาดที่อยู่อาศัยจะใช้คำจาก Internet Search Data คือคำว่า Estate agents ในแบบจำลองร่วมกับการใช้ตัวแปรอื่น โดยพบว่า สามารถอธิบายแบบจำลองดังกล่าวได้ดีขึ้น กล่าวคือทั้ง 2 แบบจำลองมี Goodness of Fit ที่ดี ซึ่งพิจารณาจากค่า Adjusted R-Square ที่สูงขึ้น ขณะที่ AIC ต่ำลง
Choi and Varian (2012)	Predicting the Present with Google Trends	วัตถุประสงค์: เพื่อใช้ Google Trends เป็นอีกหนึ่งตัวแปรในการพยากรณ์ และเพื่อทำให้แบบจำลองในการพยากรณ์แม่นยำยิ่งขึ้น วิธีการศึกษา: สร้างแบบจำลองโดยใช้ตัวแปรที่มีอยู่ และใช้ตัวแปรจาก Google Trends ที่เกี่ยวข้องกับแบบจำลองและมีความถี่ของการค้นหาเป็นรายวันหรือรายสัปดาห์ รวมกับความถี่ของ

Ref. code: 25646304010116LCN

ตารางที่ 2.1

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวชี้วัดทางเศรษฐกิจ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
		ตัวแปรที่ใช้ชี้วัดต่าง ๆ ที่มีความถี่เป็นรายไตรมาส โดยมีการศึกษาหลายแบบจำลอง ได้แก่ อัตราการว่างงาน (Unemployment Rate) จำนวนการขอรับสวัสดิการการว่างงาน (Initial Claims) ดัชนีความเชื่อมั่นของผู้บริโภค (Consumer Sentiment Index) ผลการศึกษา: พบว่า แบบจำลองส่วนใหญ่เมื่อมีการนำคำค้นหา Google Trends มาเป็นตัวแปรร่วมในการพยากรณ์มักจะมีตัวแปรตามมีความสัมพันธ์กับตัวแปรของคำค้นหาจาก Google Trends มีส่วนเบี่ยงเบนมาตรฐานลดลง ค่า log likelihood และค่า AIC ลดลง
Stephens-Davidowitz and Varian (2015)	A Hands on Guide to Google Data	วัตถุประสงค์: เพื่อใช้เครื่องมือ Google Correlate, Google Trends และ Google Consumer Survey วิธีการศึกษา: ตัวอย่างการศึกษาที่สำคัญ ได้แก่ (1) การพยากรณ์ยอดขายบ้านโดยใช้ Google Correlate ในการคาดการณ์ Out-of-sample และใช้แบบจำลอง Autoregressive Regression 1 (AR1) เปรียบเทียบกับแบบจำลองที่มีการใช้ข้อมูลของคำค้นหาใน Google searches (2) การพยากรณ์อัตราการว่างงาน โดยเปรียบเทียบระหว่างแบบจำลอง AR1 กับแบบจำลองที่มีการใช้ข้อมูลของคำค้นหาใน Google searches ผลการศึกษา: สำหรับแบบจำลองการพยากรณ์ยอดขายบ้าน พบว่า แบบจำลองที่มีการใช้ Google data มีค่าคาดเคลื่อนต่ำกว่าแบบจำลอง AR1 ส่วนแบบจำลองการพยากรณ์อัตราการว่างงาน พบว่า วิธีที่มีการใช้ข้อมูล Google data จะมีค่านัยสำคัญทางสถิติ ต่ำกว่า

หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

2.2.2 งานวิจัยที่เกี่ยวข้องที่มีการนำ Google Trends มาประยุกต์ใช้กับตัวแปรอื่น ๆ

ตารางที่ 2.2

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวแปรอื่น ๆ

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
Tantaopas (2016)	Attention effect via internet search intensity in Asia-Oceania stock markets	วัตถุประสงค์: เพื่อศึกษาความสนใจของนักลงทุนผ่านคำค้นหาในอินเตอร์เน็ตของตลาดหุ้น Asia-Oceania วิธีการศึกษา: เก็บข้อมูล Search Volume Index (SVI) เพื่อใช้ในการวัดระดับความสนใจของนักลงทุนในแต่ละกลุ่มประเทศโดยแบ่งเป็นกลุ่มประเทศที่พัฒนาแล้ว และกลุ่มประเทศที่กำลังพัฒนา และศึกษาความสัมพันธ์ของ Search Volume Index (SVI) กับตัวแปรในตลาด 3 ตัว คือ ผลตอบแทน ความผันผวน และปริมาณการซื้อ โดยใช้วิธี Vector autoregressive (VAR) และ Pairwise Granger causality ผลการศึกษา: พบว่า การเปลี่ยนแปลงของตัวแปรในตลาดการเงิน ทำให้ความสนใจของนักลงทุนที่วัดจาก SVI เปลี่ยนไปตาม
ธนสิน ถนอมพงษ์พันธ์ (2563)	การพยากรณ์ยอดขายรถยนต์ในประเทศไทยด้วยคำค้นหาใน Google	วัตถุประสงค์: เพื่อพยากรณ์ยอดขายรถยนต์จากคำค้นหาใน Google วิธีการศึกษา: เก็บข้อมูลจำนวนรถยนต์ใหม่จดทะเบียนรายเดือนจำนวน 120 ข้อมูล ของบริษัทรถยนต์ที่มีจำนวนยอดขายรถยนต์สะสมสูงสุด 5 อันดับแรก และเก็บข้อมูลคำค้นหา Google Search Word ที่สะท้อนถึงความสนใจซื้อรถยนต์ เช่น โตโยต้า ฮอนด้า จากนั้นใช้เครื่องมือ Principal Component Analysis (PCA) เพื่อจำแนก Component ที่เหมาะสม

Ref. code: 25646304010116LCN

ตารางที่ 2.2

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวแปรอื่น ๆ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
		นำ Component ที่ได้ไปสร้างแบบจำลอง Autoregressive Regression (AR) เพื่อใช้พยากรณ์ยอดขายรถยนต์ ผลการศึกษา: พบว่า คำที่มีความสัมพันธ์ระหว่างคำค้นหาและจำนวนรถยนต์ใหม่จดทะเบียน และมีความสัมพันธ์ในทิศทางเดียวกันประกอบด้วย ชื่อบริษัทจำนวน 4 บริษัท ได้แก่ โตโยต้า ฮอนด้า มิตซูบิชิ และ มาสด้า ชื่อรุ่นรถยนต์ ได้แก่ วีออส อัลติส Altis City ไทรทัน มาสด้า2 มาสด้า3 ปัญหารถยนต์ ได้แก่ Pantip รีวิว ซ่อม ปัญหา และด้านราคาและโปรโมชั่น ส่วนการพยากรณ์ยอดขายรถยนต์ พบว่า ยอดจดทะเบียนรถยนต์ใหม่มีลักษณะขึ้นลงตามฤดูกาล
อรพิชา สังขมณีนาคร (2561)	การใช้ Google Trends ในการพยากรณ์ตลาดเงิน (SET50)	วัตถุประสงค์: หาความสัมพันธ์ของคำค้นหาใน Google กับดัชนี SET50 ในตลาดเงินของประเทศไทย ว่ามีผลต่อความผันผวน ปริมาณการซื้อขายหุ้น และผลตอบแทนของหุ้นหรือไม่ วิธีการศึกษา: เก็บข้อมูล Google Trends รายสัปดาห์ของชื่อหุ้นรายตัวที่อยู่ในตลาด SET50 และเก็บข้อมูลราคาปิด ปริมาณการซื้อขายรายวัน โดยใช้ข้อมูลจำนวน 72 ข้อมูล จากนั้นคำนวณหาผลตอบแทนของดัชนี SET50 ความผันผวน (Volatility) และปริมาณการซื้อในแต่ละสัปดาห์ และนำข้อมูลที่คำนวณได้มาทำการเลือก lag p โดยการรัน Lag Order Selection Test เพื่อหา Pre-estimation for lags แล้วนำ Vector autoregressive (VAR) Model มาหาความสัมพันธ์ของข้อมูลทั้งหมด ส่วนหุ้นแต่ละตัวจะถูกทดสอบด้วย Multiple Regression Model โดยจะมีแบบจำลองเพื่อการอธิบาย (Explanatory Model) และแบบจำลองเพื่อการพยากรณ์ (Predictive Model)

Ref. code: 25646304010116LCN

ตารางที่ 2.2

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องด้านตัวแปรอื่น ๆ (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
		ผลการศึกษา: จากการทดสอบ VAR พบว่า ไม่มีความสัมพันธ์กันทั้งคำค้นหาใน Google ความผันผวน และปริมาณการซื้อขาย ส่วนการทดสอบความสัมพันธ์ของหุ้นรายตัวจากวิธี Multiple Regression Model พบว่า มีความสัมพันธ์กับปริมาณซื้อขายหุ้นมากกว่าครึ่งหนึ่งของจำนวนหุ้นที่นำมาทดสอบ และการทดสอบความสัมพันธ์ของ คำค้นหาใน Google กับผลตอบแทนของหุ้นรายตัวใน SET50 กลับพบว่าส่วนใหญ่ไม่มีความสัมพันธ์กัน และไม่สามารถพยากรณ์หรือคาดการณ์แนวโน้มของผลตอบแทนได้
Preis, Moat and Stanley (2013)	Quantifying Trading Behavior in Financial Markets Using Google Trends	วัตถุประสงค์: เพื่อศึกษาความสัมพันธ์ของพฤติกรรมของมนุษย์ผ่าน Internet ที่ส่งผลต่อตลาดการเงิน และเพื่อพยากรณ์สัญญาณการเคลื่อนไหวของตลาดเงิน วิธีการศึกษา: วิเคราะห์ข้อมูลจากคำค้นหาจำนวน 98 คำ ที่เกี่ยวกับตลาดหุ้น เปรียบเทียบผลการดำเนินงานสะสมของกลยุทธ์ buy and hold กับ กลยุทธ์ Google Trends ผลการศึกษา: ข้อมูล Google Trends สามารถสะท้อนให้เห็นถึงเศรษฐกิจปัจจุบัน และแนวโน้มในอนาคต ซึ่งสามารถอธิบายได้ผ่านพฤติกรรมของมนุษย์ในการใช้ Internet หาข้อมูลก่อนการทำพฤติกรรมใด ๆ โดยจากการศึกษาพบว่า ปริมาณของคำค้นหาใน Google Trends ที่เกี่ยวกับตลาดเงินจะเพิ่มขึ้นก่อนที่ตลาดหุ้นจะตก และการใช้กลยุทธ์ Google Trends สามารถสร้างกำไรในตลาดหุ้นได้มากกว่า กลยุทธ์ buy and hold

หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

2.2.3 งานวิจัยที่เกี่ยวข้องที่มีการนำแบบจำลองด้านปัญญาประดิษฐ์มาประยุกต์ใช้ประมาณการตัวแปรต่าง ๆ

ตารางที่ 2.3

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องที่ประยุกต์ใช้แบบจำลอง Machine Learning

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
กานต์สินี เจริญกิจวัชรชัย (2561)	ปัญญาประดิษฐ์ กับการพยากรณ์ค่าจ้างแรงงานไทย	วัตถุประสงค์: เพื่อสร้างแบบจำลองกำหนดค่าจ้างแรงงานจากปัจจัยต่าง ๆ ด้วยวิธีการทางเศรษฐมิติ และปัญญาประดิษฐ์ และเปรียบเทียบประสิทธิภาพของแบบจำลอง วิธีการศึกษา: แบ่งข้อมูลที่ใช้ศึกษาออกเป็น Training Data และ Testing Data ในสัดส่วน 70:30 จากนั้นนำมาสร้างแบบจำลองด้วยวิธี Multiple Linear Regression, Neural network และ Random Forest ผลการศึกษา: พบว่า แบบจำลองที่ถูกสร้างด้วยวิธีการทางปัญญาประดิษฐ์ มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าวิธีการทางเศรษฐมิติเสมอ
ฤทธิไกร ทวีเจริญ (2559)	ตัวแบบการพยากรณ์การลาออกของพนักงานฝ่ายผลิตในธุรกิจเครื่องฉีดพลาสติกของบริษัทแห่งหนึ่ง	วัตถุประสงค์: เพื่อหาปัจจัยที่มีอิทธิพลกับการลาออกและพยากรณ์พนักงานที่มีแนวโน้มลาออก วิธีการศึกษา: แบ่งข้อมูลที่ใช้ศึกษาออกเป็น Training Data และ Testing Data ในสัดส่วน 70:30 และใช้เทคนิคการเรียนรู้เพื่อสร้างแบบจำลองด้วยเครื่อง Machine Learning ได้แก่ Decision Tree และ Decision Forest ผลการศึกษา: พบว่า ปัจจัยของข้อมูลประชากรศาสตร์ด้านอายุ สถานะสมรส และการศึกษา ไม่พบอิทธิพลที่ส่งผลกับการลาออก ในขณะที่เพศมีอิทธิพลที่ส่งผลกับการลาออก

Ref. code: 25646304010116LCN

ตารางที่ 2.3

การสรุปผลการทบทวนวรรณกรรมปริทัศน์และงานวิจัยที่เกี่ยวข้องที่ประยุกต์ใช้แบบจำลอง Machine Learning (ต่อ)

ผู้ศึกษา (ปีการศึกษา)	เรื่องที่ศึกษา	การวิเคราะห์ข้อมูลและผลการศึกษา
จิราภรณ์ เจริญยิ่ง (2563)	การพยากรณ์ผลสัมฤทธิ์ทางการเรียน ด้วยเทคนิคเหมืองข้อมูล โดยใช้ Rapid Miner	วัตถุประสงค์: เพื่อศึกษาวิธีที่เหมาะสมในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยการทำเหมืองข้อมูล วิธีการศึกษา: แบ่งข้อมูลที่ใช้ศึกษาออกเป็น Training Data และ Testing Data ในสัดส่วน 80:20 และสร้างแบบจำลองการพยากรณ์ด้วยการทำเหมืองข้อมูล (Data Mining) ด้วยเทคนิค (Decision Tree) เทคนิคป่าแห่งการทำนาย (Random Forest) การเรียนรู้เบย์ (Naïve-Bayes) และ K-NN (K-Nearest Neighbors) ผลการศึกษา: พบว่า ค่าความถูกต้องของการพยากรณ์ (accuracy) ด้วยเทคนิค Random Forest เท่ากับร้อยละ 100 ซึ่งมากที่สุดเมื่อเทียบกับเทคนิคอื่น
Puttanapong, Martinez Jr., Addawe, Bulan, et al. (2020)	Predicting Poverty Using Geospatial Data in Thailand	วัตถุประสงค์: เพื่อใช้ข้อมูลจากดาวเทียบและ Machine Learning ในการพยากรณ์ความยากจนแต่ละจังหวัดในประเทศไทย วิธีการศึกษา: สร้างแบบจำลองการพยากรณ์ด้วยวิธี generalized least squares (GLS) neural network (NN), random forest (RF) และ support vector regression (SVR) ผลการศึกษา: พบว่า RF ให้ผลการพยากรณ์ที่แม่นยำที่สุดในการพยากรณ์ความยากจนของแต่ละจังหวัด โดยมีความแม่นยำร้อยละ 80

หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

บทที่ 3
วิธีการวิจัย

3.1 ข้อมูลที่ใช้ในการศึกษา

3.1.1 แหล่งข้อมูล

ข้อมูลที่ใช้สำหรับการศึกษานี้เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีลักษณะเป็นข้อมูลทุติยภูมิ (Secondary Data) ที่อยู่ในรูปแบบของอนุกรมเวลา (Time Series Data) ที่มีความถี่ของข้อมูลแบบรายเดือน ซึ่งรวบรวมข้อมูลจากหน่วยงานที่เกี่ยวข้อง โดยสามารถแบ่งได้เป็นตัวแปรต้น และตัวแปรตามได้ ดังนี้

3.1.1.1 ข้อมูลตัวแปรต้น หรือแปรอิสระ (Independent Variable)

ทำการศึกษาจากการเก็บข้อมูลจาก Google Trends ที่มีแหล่งที่มาจากฐานข้อมูลของ Google ณ วันที่ 30 กันยายน พ.ศ. 2564 ซึ่งเป็นข้อมูลที่แสดงในลักษณะความถี่เป็นรายเดือน และเป็นข้อมูลเปรียบเทียบกับคำค้นหาอื่น ๆ ที่แสดงเป็นสัดส่วนความสนใจจาก 0 - 100 ซึ่งข้อมูล Google Trends นี้จะเป็นข้อมูลจากคำค้นหาที่เกี่ยวข้องกับการหางานและอยู่ในประเทศไทย โดยจะใช้กลุ่มคำจำนวน 24 คำ[1] ได้แก่ สมัครงาน, หางาน, ว่างงาน, ประกันสังคม, สัมภาษณ์งาน, นายจ้างหาคนงาน, งานราชการ, สมัครงานบริษัท, หางานสมัครงาน, ใบสมัครงาน, หางานทำ, สิทธิประกันสังคม, ตกงาน, Jobs DB, Job Top Gun, LinkedIn, Job Thai, Job TH, กรมการจัดหางาน, Job Fair, เงินชดเชยว่างงาน, ลงทะเบียนว่างงาน, กองทุนสำรองเลี้ยงชีพ และ เงินชดเชยเลิกจ้าง

3.1.1.2 ข้อมูลตัวแปรตาม (Dependent Variable)

การศึกษานี้จะประกอบด้วย 2 ชุดข้อมูลจาก 2 แหล่งที่มาของข้อมูล ดังนี้

(1) ข้อมูลอัตราการว่างงาน (Unemployment Rate) จากการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (National Statistical Office: NSO)[2] ที่มีความถี่เป็นรายเดือนซึ่งเป็นการเก็บข้อมูล และจะใช้ข้อมูลรายเดือนเป็นระยะเวลา 10 ปี ย้อนหลัง ระหว่างปี พ.ศ. 2554 - 2563 โดยในปี พ.ศ. 2563 ช่วงเดือนเมษายน - มิถุนายน จะไม่มีการเก็บข้อมูลเนื่องจากการ

---

[1] มีที่มาของข้อมูลจากเว็บไซต์ https://trends.google.com/trends

[2] มีที่มาของข้อมูลจากเว็บไซต์ www.bot.or.th/App/BIZSHR/stat/Graph/ ของธนาคารแห่งประเทศไทย

Ref. code: 25646304010116LCN

ระบาดโควิด-19 รุนแรง สำนักงานสถิติจึงไม่ได้จัดส่งเจ้าหน้าที่ออกสำรวจ จึงทำให้ในการศึกษาส่วนนี้จะแบ่งข้อมูลออกเป็น 2 ส่วน คือข้อมูลก่อนเดือนเมษายน 2563 และข้อมูลหลังมิถุนายน 2563

(2) ข้อมูลอัตราการว่างงาน (Unemployment Rate) จากผู้ประกันตนในประกันสังคม (Social Security Office: SSO)[3] ที่ขอรับเงินชดเชยกรณีว่างงานต่อจำนวนผู้ประกันตนในประกันสังคมตามมาตรา 33, 39 และ 40 ที่มีความถี่เป็นรายเดือน ซึ่งเป็นสัดส่วนของข้อมูล โดยใช้ข้อมูลรายเดือนตั้งแต่เดือนเป็นระยะเวลา 10 ปี ย้อนหลัง ระหว่างมกราคม พ.ศ. 2554 – พฤศจิกายน พ.ศ. 2564

3.1.2 รายละเอียดของข้อมูลที่ใช้ในการศึกษา

เมื่อได้ตัวแปรจากข้อ 3.1.1 มาใส่ในแบบจำลองอัตราการว่างงานไทยกับ Google Trends โดยใช้แบบจำลอง 5 วิธี คือ แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model) แบบจำลองต้นไม้ตัดสินใจ (Decision Tree Regression Model) แบบจำลองซับพอรตเวกเตอรรีเกรสชัน (Support Vector Regression Model: SVR) แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN) และแบบจำลองแบบเรียงซ้อน (Ensemble Model: EM) จากนั้นจะเปรียบเทียบผลลัพธ์แต่ละแบบจำลองเพื่อวิเคราะห์ความแม่นยำของการพยากรณ์อัตราการว่างงานไทยในแต่ละชุดข้อมูล

ข้อมูลที่ใช้ในการศึกษาที่เป็นข้อมูลระหว่างมกราคม พ.ศ. 2554 – พฤศจิกายน พ.ศ. 2564 จะพบว่าชุดข้อมูลจากการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (National Statistical Office: NSO) มีจำนวนข้อมูล 117 ตัวอย่าง และชุดข้อมูลจากผู้ประกันตนในประกันสังคม (Social Security Office: SSO) จำนวน 130 ตัวอย่าง โดยในส่วนของข้อมูลตัวแปรต้นหรือตัวแปร Google Trends ได้ตัดข้อมูลออกในบางช่วงเวลาเพื่อให้สอดคล้องกับข้อมูลอัตราการว่างงาน ทั้งนี้ สามารถอธิบายรายละเอียดลักษณะของข้อมูล 10 ปี ที่ใช้ในแบบจำลองพนยากรณ์อัตราการว่างงานได้ดังต่อไปนี้

3.1.2.1 ข้อมูล Google Trends

ข้อมูลที่ได้จากกลุ่มตัวอย่างซึ่งคัดเลือกจากคำค้นหาที่เกี่ยวข้องกับการหางาน สามารถแบ่งเป็น 4 กลุ่ม ได้แก่

[3] มีที่มาของข้อมูลจากเว็บไซต์ www.warning.mol.go.th/article/review ของสำนักงานสถิติแห่งชาติ

Ref. code: 25646304010116LCN

(1) กลุ่มคำจากงานวิจัยที่เกี่ยวข้องโดยจะเลือกคำที่มีค่าความสัมพันธ์ระหว่างอัตราการว่างงานสูงเมื่อเทียบกับตัวอื่น อาทิ สมัครงาน, หางาน, ว่างงาน, ประกันสังคม, และสัมภาษณ์งาน

(2) กลุ่มคำที่เป็นคำยอดนิยมหรือคำมาแรงที่ Google Trends จับข้อมูลได้ อาทิ นายจ้างหาคนงาน, งานราชการ, สมัครงานบริษัท, หางานสมัครงาน, ใบสมัครงาน, หางานทำ, สิทธิประกันสังคม และตกงาน

(3) กลุ่มคำที่เป็นชื่อบริษัทจัดหางาน อาทิ Jobs DB, Job Top Gun, Linked in, Job Thai, Job TH, กรมการจัดหางาน และ Job Fair

(4) กลุ่มคำที่เป็นคำที่เกี่ยวข้องกับสวัสดิการหางาน อาทิ เงินชดเชยว่างงาน, ลงทะเบียนว่างงาน, กองทุนสำรองเลี้ยงชีพ และเงินชดเชยเลิกจ้าง

ทั้งนี้ ข้อมูลของกลุ่มคำที่กล่าวมาข้างต้นจะแสดงข้อมูลเชิงปริมาณ พบว่า คำค้นหา “งานราชการ” จะเป็นคำที่มีสัดส่วนการค้นหามากที่สุดในช่วงปี พ.ศ. 2554 – 2561 และเริ่มลดลงหลังจากปีดังกล่าว ขณะที่คำว่า “สมัครงาน” เป็นคำที่มีสัดส่วนการค้นหาค่อนข้างทรงตัว อย่างไรก็ตามคำว่า “ประกันสังคม” และ “ลงทะเบียนว่างงาน” เป็นคำที่มีสัดส่วนการค้นหามากในปี พ.ศ. 2563 ในขณะที่ปีอื่น ๆ มีสัดส่วนการค้นหาต่ำ เนื่องจากในปี พ.ศ. 2563 เป็นปีที่มีสถานการณ์การแพร่ระบาดของเชื้อโควิด – 19 ทำให้หน่วยงานภาครัฐต้องเข้ามามีบทบาทในการช่วยเหลือผู้ที่ว่างงานจากสถานการณ์ดังกล่าว

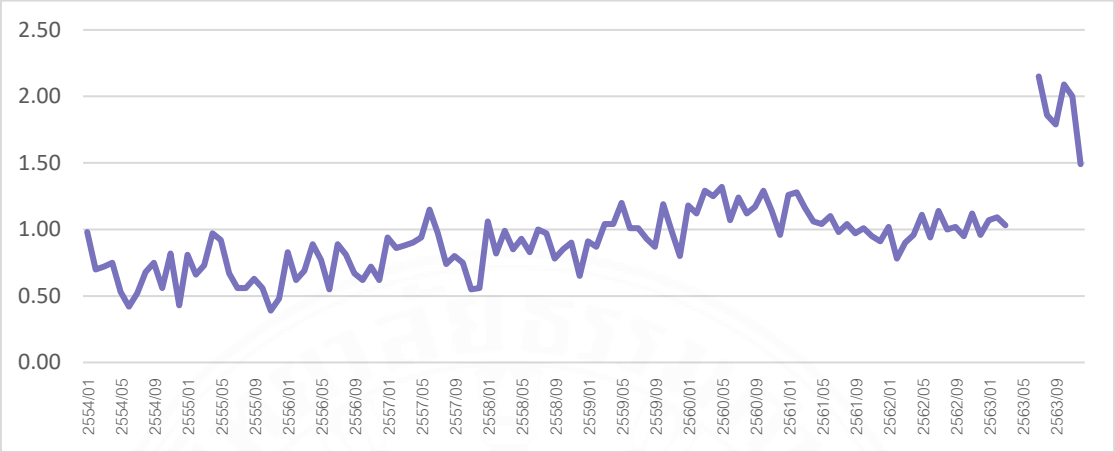
3.1.2.2 ข้อมูลอัตราการว่างงานจากสำนักงานสถิติแห่งชาติ (NSO)

จากการรวบรวมข้อมูลอัตราการว่างงานจากการสำรวจประชากรของสำนักงานสถิติแห่งชาติ พบว่า อัตราการว่างงานรายเดือนตั้งแต่ปี พ.ศ. 2554 จะอยู่ในช่วงระหว่างร้อยละ 0.39 – 1.00 ในขณะที่บางเดือนอัตราการว่างงานจะเกินร้อย 1 อย่างไรก็ตาม ในช่วงปี พ.ศ. 2559 อัตราการว่างงานเริ่มขยับขึ้นมาเกินกว่าร้อยละ 1 จนกระทั่งในปี พ.ศ. 2563 ได้เกิดสถานการณ์การแพร่ระบาดของเชื้อโควิด – 19 ส่งผลให้ไม่มีการการสำรวจประชากรของสำนักงานสถิติแห่งชาติในเดือนเมษายน – มิถุนายน พ.ศ. 2563 จากนั้นเมื่อมีการสำรวจอัตราการว่างงานอีกครั้งกลับพบว่าอัตราการว่างงานเพิ่มขึ้นถึงประมาณร้อยละ 2 เป็นผลมาจากกิจกรรมทางเศรษฐกิจที่หยุดชะงักบริษัทหลายแห่งปลดแรงงาน โดยจะแสดงข้อมูลดังภาพที่ 3.1

Ref. code: 25646304010116LCN

ภาพที่ 3.1

อัตราการว่างงานรายเดือนในประเทศไทยตั้งแต่ปี พ.ศ. 2554 – 2563



หมายเหตุ. โดยสำนักงานสถิติแห่งชาติ, 2564, (https://www.bot.or.th/App/BIZSHR/stat/Graph/)

3.1.2.3 ข้อมูลอัตราการว่างงานจากผู้ประกันตนในประกันสังคม (SSO)

เป็นการรวบรวมข้อมูลจากสำนักงานประกันสังคม โดยใช้ข้อมูลของผู้ประกันตนในประกันสังคม ซึ่งประกอบไปด้วยผู้ประกันตนที่รับประโยชน์ทดแทนกรณีการว่างงาน ผู้ประกันตนภาคสมัครใจตามมาตรา 39 และมาตรา 40 และผู้ประกันตนภาคบังคับตามมาตรา 33 โดยสามารถเขียนวิธีการคำนวณได้ ดังนี้

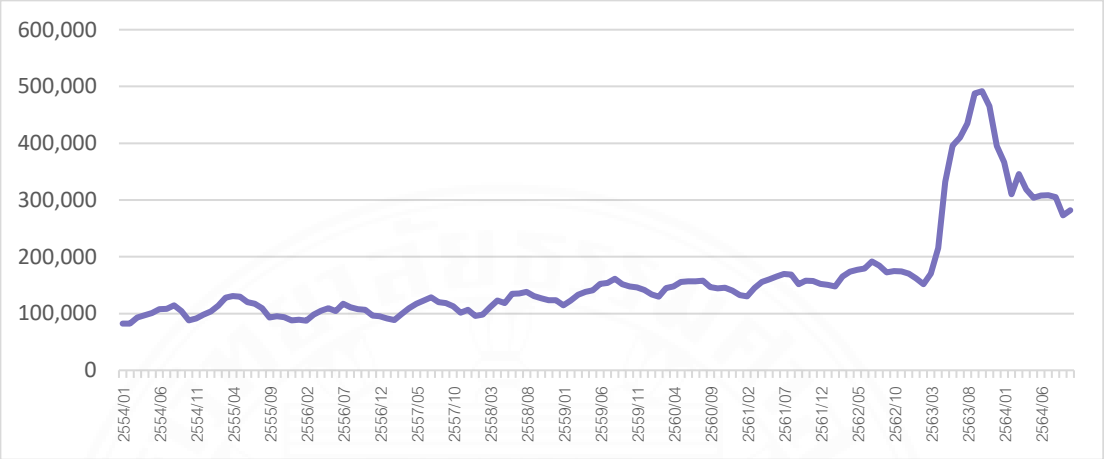
$$\text{อัตราการว่างงาน} = \frac{\text{ผู้ประกันตนที่รับประโยชน์ทดแทนกรณีว่างงาน}}{\text{ผู้ประกันตนตามมาตรา 33 มาตรา 39 และมาตรา 40}}$$

ทั้งนี้ พบว่า ข้อมูลอัตราการว่างงานจากสำนักงานประกันสังคมตั้งแต่ปี พ.ศ. 2554 อยู่ในช่วงที่ไม่ค่อยแตกต่างกันมากในแต่ละเดือน อย่างไรก็ตาม ในช่วงเดือนเมษายน พ.ศ. 2563 – พฤศจิกายน พ.ศ. 2564 อัตราการว่างงานเริ่มเพิ่มขึ้นอยู่ในช่วงระหว่างร้อยละ 1.23 – 3.50 เป็นผลมาจากสถานการณ์การแพร่ระบาดของเชื้อไรวัสโควิด – 19 ที่ทำให้ผู้ประกันตนในประกันสังคมว่างงานเพิ่มขึ้น โดยจะแสดงข้อมูลดังภาพที่ 3.2

Ref. code: 25646304010116LCN

ภาพที่ 3.2

อัตราการว่างงานของผู้ประกันตนที่รับประโยชน์ทดแทนกรณีการว่างงานรายเดือนของสำนักงานประกันสังคมตั้งแต่ มกราคม พ.ศ. 2554 – ตุลาคม พ.ศ. 2564



หมายเหตุ. โดยสำนักงานประกันสังคม, 2564, (http://warning.mol.go.th/uploadFile/pdf/pdf-2021-11-24-1637721494.pdf)

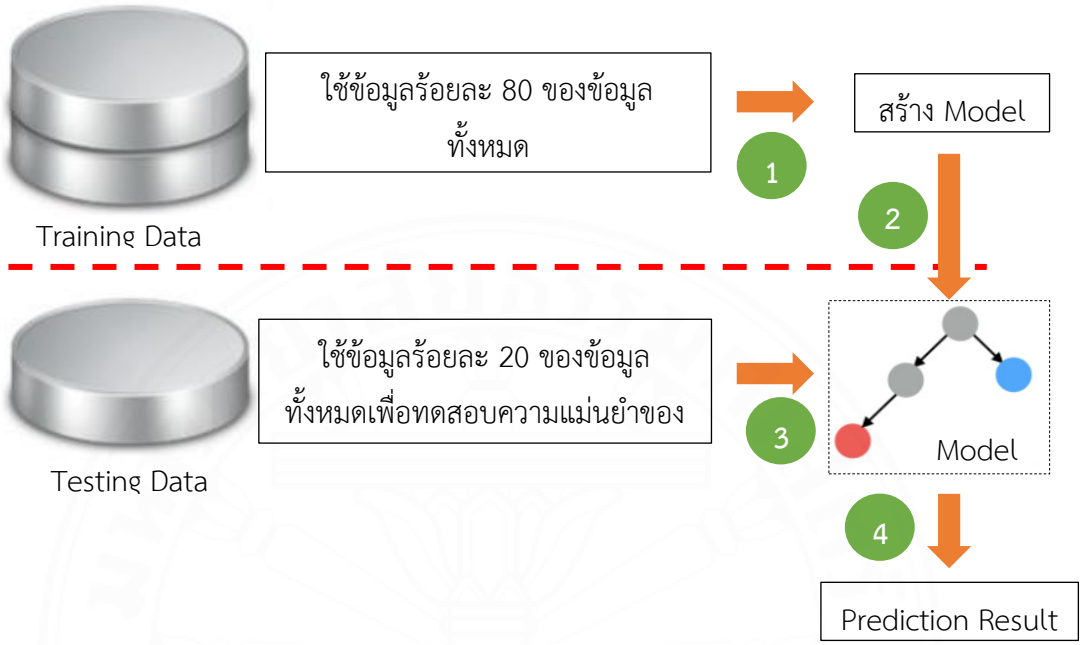
3.1.3 การแบ่งชุดข้อมูลเพื่อทดสอบประสิทธิภาพของแบบจำลอง

ในการศึกษานี้จะแบ่งข้อมูลสำหรับนำไปใช้ศึกษาในแต่ละแบบจำลองและเพื่อใช้ทดสอบประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยวิธี Split Test โดยจะแบ่งข้อมูลเป็นข้อมูลที่ใช้ในการเรียนรู้เพื่อสร้างแบบจำลอง (Training Data) ร้อยละ 80 และข้อมูลที่ใช้ทดสอบความแม่นยำหรือประสิทธิภาพของแบบจำลอง (Testing Data) ร้อยละ 20 ซึ่งในการแบ่งข้อมูลนี้จะนำชุดข้อมูลแต่ละชุด ประกับด้วยชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) และชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) นำมาเรียงตามข้อมูลรายเดือนในช่วงเวลาเดียวกัน ดังภาพที่ 3.3

Ref. code: 25646304010116LCN

ภาพที่ 3.3

การแบ่งข้อมูลออกเป็น Training Data และ Testing Data



หมายเหตุ. รวบรวมโดยผู้ศึกษา

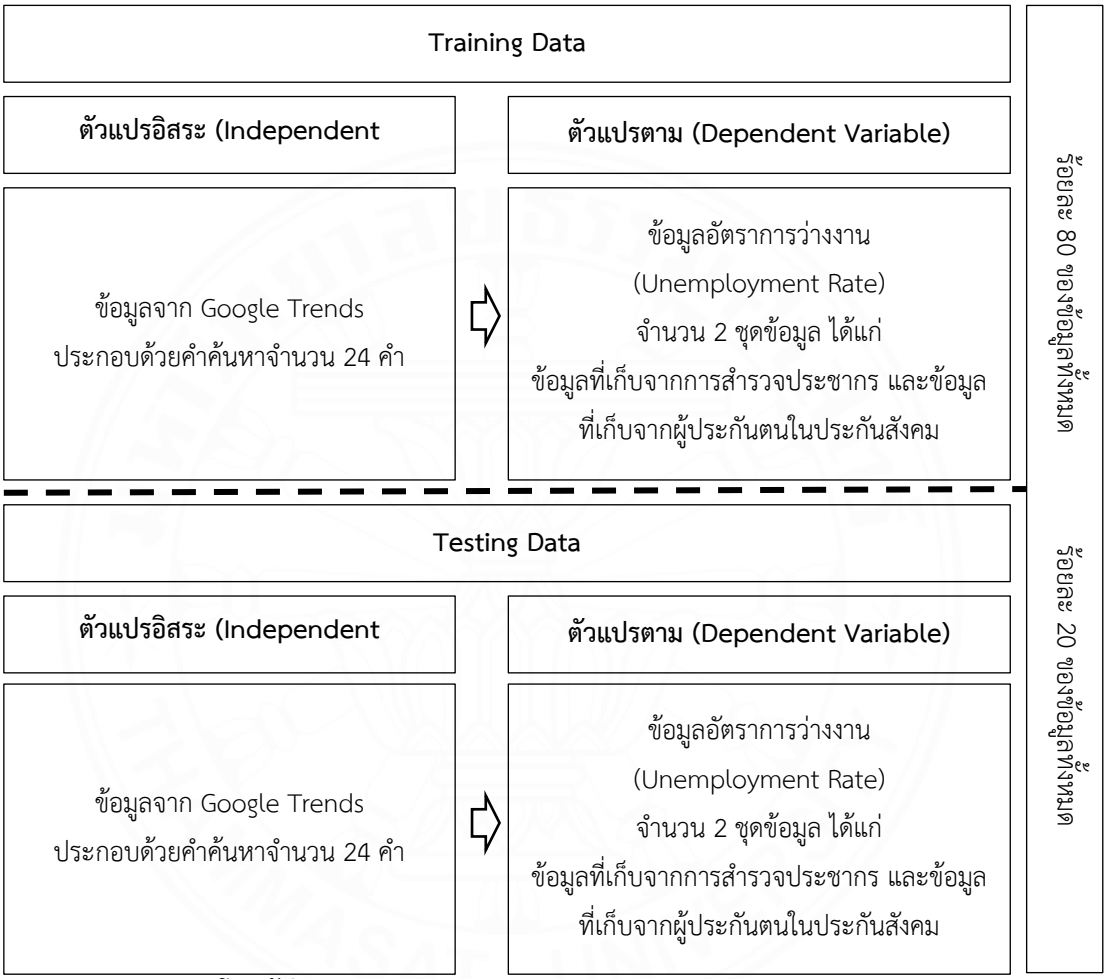
วิธีการทดสอบประสิทธิภาพของแบบจำลองด้วยวิธี Split Test โดยจะแบ่งข้อมูลออกเป็น 2 ส่วน คือร้อยละ 80 ต่อร้อยละ 20 โดยข้อมูลส่วนแรกร้อยละ 80 จะใช้ในการสร้างแบบจำลอง (Training Data) เพื่อให้แบบจำลองของ Machine Learning สามารถเรียนรู้ได้จากข้อมูลส่วนนี้ และข้อมูลส่วนที่สองร้อยละ 20 ใช้ในการทดสอบความแม่นยำหรือประสิทธิภาพของแบบจำลอง (Testing Data) ดังภาพที่ 3.3 โดยการทดสอบแบบ Split Test จะทำการนำข้อมูล Testing Data มาทำการเปรียบเทียบกับผลลัพธ์ของแบบจำลองที่ถูกสร้างด้วย Training Data ว่าวิธีการของแบบจำลองใด ได้ผลผลัพธ์ที่แม่นยำที่สุด ซึ่งข้อดีของวิธีการนี้คือใช้เวลาในการสร้างแบบจำลองน้อยลง และเหมาะกับชุดข้อมูลที่มีขนาดใหญ่

Ref. code: 25646304010116LCN

3.2 กรอบแนวคิดในการศึกษา

ภาพที่ 3.4

กรอบแนวคิดในการศึกษา



หมายเหตุ. รวบรวมโดยผู้ศึกษา.

3.3 แบบจำลองที่ใช้ในการศึกษา

ในส่วนของงานศึกษานี้จะใช้แบบจำลองการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) ที่มีลักษณะเป็นอนุกรมเวลา (Time Series Data) กล่าวคือ เป็นชุดของข้อมูลที่มีการเก็บรวบรวมตามระยะเวลาที่ติดต่อกันอย่างเป็นระบบ และใช้เครื่องมือ Machine Learning (ML) ซึ่งเป็นศาสตร์การเรียนรู้ของ Artificial Intelligence (AI) ในการวิเคราะห์ข้อมูล โดยงานศึกษานี้จะใช้โปรแกรม MATLAB ในการสร้างแบบจำลอง ได้แก่ แบบจำลองการถดถอยเชิงเส้น (Linear

Ref. code: 25646304010116LCN

Regression Model), แบบจำลองต้นไม้ตัดสินใจ (Decision Tree Regression Model), แบบจำลองซับพอรตเวกเตอรรีเกรสชัน (Support Vector Regression Model: SVR), แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN) และแบบจำลองแบบเรียงซ้อน (Ensemble Model: EM) ซึ่งจะอธิบายรายละเอียดของวิธีการศึกษา ดังนี้

3.3.1 แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model)

เป็นการวิเคราะห์ความถดถอยเชิงเส้นเพื่อแสดงการความสัมพันธ์ระหว่างตัวแปรตาม (Y) และตัวแปรอิสระ (X) ของประชากรจะเห็นว่า กลุ่มตัวแปรอิสระ $(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$ สามารถอธิบายการเปลี่ยนแปลงค่าของตัวแปรตามได้ส่วนหนึ่ง ในส่วนของการเปลี่ยนแปลงที่ไม่สามารถอธิบายได้นี้ เรียกว่า ค่าความคาดเคลื่อนในการพยากรณ์ (Error: ε) การวิเคราะห์เชิงถดถอยแบบพหุคูณจะเป็นการพยากรณ์หาค่าสัมประสิทธิ์ α และ β_1 จากค่าสถิติ a และ b ที่ได้จากการคำนวณโดยกลุ่มตัวอย่าง โดยหลักการวิเคราะห์ คือ ค่าสัมประสิทธิ์ที่คำนวณได้จะต้องเป็นค่าสัมประสิทธิ์ที่ทำให้สมการดังกล่าวมีค่าความคาดเคลื่อนกำลังสองรวมกันน้อยที่สุด (Ordinary Least Square: OLS) (อัปษรศรี ม่วงคง, 2552)

ในงานศึกษานี้จะเป็นการสร้างแบบจำลองเพื่อพยากรณ์อัตราการว่างงานไทย โดยใช้ข้อมูล Google Trends เป็นตัวแปรในการพยากรณ์ ซึ่งจะแสดงความสัมพันธ์ระหว่างอัตราการว่างงานและ Google Trends สามารถอธิบายได้ ดังนี้

$$\widehat{\text{Unemployment}}\ \text{Rate} = f(x_1, x_2, \ldots, x_{24})$$

โดยที่ $\widehat{\text{Unemployment}}\ \text{Rate}$ คือ อัตราการว่างงานจาก NSO หรือ SSO

$x_1, x_2, \ldots, x_{24}$ คือ ข้อมูลของคำค้นหาเกี่ยวกับการหางานจาก Google Trends

จากนั้นจะประมาณความสัมพันธ์ดังกล่าวเป็นแบบเชิงเส้นและใช้สมการถดถอยเชิงเส้นเพื่อวิเคราะห์ความสัมพันธ์ของแต่ละตัวแปรในสมการ ดังนี้

Ref. code: 25646304010116LCN

Unemployment Rate

$$= a_1 + \beta_1 \text{Apply for a company job} + \beta_2 \text{Apply for work} + \beta_3 \text{Available} + \beta_4 \text{Compensation} + \beta_5 \text{DOE} + \beta_6 \text{Employers looking for workers} + \beta_7 \text{Find a job} + \beta_8 \text{Find a job apply for a job} + \beta_9 \text{Find work} + \beta_{10} \text{Government job} + \beta_{11} \text{Interview} + \beta_{12} \text{Job application} + \beta_{13} \text{Jobfair} + \beta_{14} \text{Jobsdb} + \beta_{15} \text{Jobth} + \beta_{16} \text{Jobthai} + \beta_{17} \text{Jobtopgun} + \beta_{18} \text{Linkedin} + \beta_{19} \text{Provident fund} + \beta_{20} \text{Social Security} + \beta_{21} \text{Social security rights} + \beta_{22} \text{Unemploy} + \beta_{23} \text{Unemployment compensation} + \beta_{24} \text{Unemployment registration} + \varepsilon_i$$

โดยที่ a_1 คือ ค่าคงที่

β_1 คือ สัมประสิทธิ์ของสมัครงานบริษัท

β_2 คือ สัมประสิทธิ์ของหางาน

β_3 คือ สัมประสิทธิ์ของว่างงาน

β_4 คือ สัมประสิทธิ์ของเงินชดเชยเลิกจ้าง

β_5 คือ สัมประสิทธิ์ของกรมการจัดหางาน

β_6 คือ สัมประสิทธิ์ของนายจ้างหาคนงาน

β_7 คือ สัมประสิทธิ์ของสมัครงาน

β_8 คือ สัมประสิทธิ์ของหางานสมัครงาน

β_9 คือ สัมประสิทธิ์ของหางานทำ

β_{10} คือ สัมประสิทธิ์ของงานราชการ

β_{11} คือ สัมประสิทธิ์ของสัมภาษณ์งาน

β_{12} คือ สัมประสิทธิ์ของใบสมัครงาน

β_{13} คือ สัมประสิทธิ์ของ JobFairไทย

β_{14} คือ สัมประสิทธิ์ของ JobsDBไทย

β_{15} คือ สัมประสิทธิ์ของ JobTHไทย

β_{16} คือ สัมประสิทธิ์ของ JobThaiไทย

β_{17} คือ สัมประสิทธิ์ของ JobTopGunไทย

β_{18} คือ สัมประสิทธิ์ของ LinkedInไทย

β_{19} คือ สัมประสิทธิ์ของกองทนสำรองเลี้ยงชีพ

β_{20} คือ สัมประสิทธิ์ของประกันสังคม

β_{21} คือ สัมประสิทธิ์ของสิทธิประกันสังคม

Ref. code: 25646304010116LCN

β_{22} คือ สัมประสิทธิ์ของตกงาน

β_{23} คือ สัมประสิทธิ์ของเงินชดเชยว่างงาน

β_{24} คือ สัมประสิทธิ์ของลงทะเบียนว่างงาน

ε_i คือ ค่าความคลาดเคลื่อน

การวิเคราะห์ Linear Regression จะเป็นการพยากรณ์หาค่าสัมประสิทธิ์ที่ได้จากการคำนวณโดยกลุ่มตัวอย่าง มีหลักการวิเคราะห์ คือ ค่าสัมประสิทธิ์ที่คำนวณได้จะต้องเป็นค่าสัมประสิทธิ์ที่ทำให้แบบจำลองดังกล่าว มีค่าความคาดเคลื่อนกำลังสองรวมกันน้อยที่สุด (Ordinary Least Square: OLS)

3.3.2 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree Regression Model)

ในงานศึกษานี้ จะเป็นการเรียนรู้ของตัวแบบจำลองจากการจำแนกประเภทข้อมูลซึ่งมีหลักการในการประมาณการอัตราการว่างงานโดยจะแบ่งข้อมูลออกทีละ 2 ส่วน จากโหนดล่างสุดของต้นไม้ที่เรียกว่าโหนดราก จากนั้นจะไล่ขึ้นมาเรื่อย ๆ จนถึงโหนดบนสุดของต้นไม้ที่เรียกว่าโหนดใบ และ สร้างโหนดและเส้นเชื่อมไปเรื่อย ๆ จนกว่าข้อมูลที่ได้จะถูกจัดไว้เป็นกลุ่มเดียวกันถึงจะหยุดสร้างต้นไม้ จากนั้นจะทำการประมาณการอัตราการว่างงานโดยใช้ค่าเฉลี่ยของค่าอัตราการว่างงาน ซึ่งจะทำการแบ่งข้อมูลจากโหนดรากไปจนถึงโหนดใบ จนกว่าจะได้ตามเงื่อนไขที่กำหนด สำหรับในการศึกษานี้จะมีหลักการในการแบ่งข้อมูลแต่ละโหนด ดังนี้

(1) จะเลือก 1 ตัวอย่างจากข้อมูล Google Trends ทั้งหมด มาทำการคัดกรองข้อมูลด้วยค่าของตัวอย่างนั้นที่เลือกมา

(2) หาจุดแบ่งที่เป็นไปได้ทั้งหมดจากจำนวนข้อมูลทั้งหมด

(3) ในการแบ่งข้อมูลแต่ละแบบที่เป็นไปได้ จะคำนวณหาค่า residual sum of squares (RSS) จากการทำประมาณการอัตราการว่างงานด้วยค่าเฉลี่ยของอัตราการว่างงานในแต่ละกลุ่ม โดยมีสูตร ดังนี้

$$\mathrm{RSS} = \sum_{\mathrm{j}=1}^{\mathrm{J}} \sum_{\mathrm{i} \in \mathrm{R_j}} (\mathrm{y_i} - \widehat{\mathrm{y_{R_j}}})^2$$

โดยที่ R_j คือ แต่ละกลุ่มของ observation ที่ถูกแบ่งออกมา เป็นทั้งหมด J กลุ่ม

y_i คือ อัตราการว่างงาน

$\widehat{y_{R_j}}$ คือ ค่าประมาณการอัตราการว่างงานในแต่ละกลุ่มที่คำนวณมาจากค่าเฉลี่ยของอัตราการว่างงานในกลุ่มนั้น ๆ

Ref. code: 25646304010116LCN

(4) จากนั้นจะทำการแยกตัวอย่างที่มีค่า RSS น้อยที่สุดออกมา เมื่อสิ้นสุดการแยกแล้ว จะประมาณการค่าอัตราการว่างงานจากค่าเฉลี่ยของอัตราการว่างงานภายในโหนดของตัวเอง สำหรับโหนดที่ยังสามารถแยกได้ต่อ โดยจะเรียกค่าสุดท้ายของการแยกที่ให้ค่าประมาณการมาเป็นค่าเฉลี่ยของอัตราการว่างงานในแต่ละกลุ่มว่าโหนดใบ

3.3.3 แบบจำลองซับพอรตเวกเตอร์รีเกรสชัน (Support Vector Regression Model: SVR)

เป็นแนวคิดพื้นฐานของแบบจำลองการพยากรณ์แบบดั้งเดิม กล่าวคือเพื่อแก้ไขปัญหาการหาค่าที่เหมาะสมที่สุด (Optimization problem) โดยเฉพาะการลดข้อผิดพลาดรวมที่อยู่ในรูปแบบเฉพาะของฟังก์ชันการสูญเสีย (Loss Function) Vapnik (1998) ได้แนะนำแนวทางใหม่เพื่อขจัดปัญหาที่กล่าวมาข้างต้น โดยการใช้ Support Vector สำหรับการแบ่งส่วนของข้อมูลในปัญหาของ Classification ในทางคณิตศาสตร์ โดยการเปลี่ยนจาก Support Vector Machine (SVM) ไปยัง Support Vector Regression (SVR) โดยมุ่งเป้าไปที่การลดปัญหาการเกิดฟังก์ชันการสูญเสียได้ในระดับหนึ่ง ทำให้แบบจำลองนี้มีความยืดหยุ่นมากขึ้นในการพยากรณ์ โดยเฉพาะการผ่อนคลายข้อจำกัดทั่วไปที่เกี่ยวข้องกับคุณสมบัติการกระจายตัวของตัวแปร (Zang et al., 2010 and Wang et al., 2012) ด้วยกรอบการวิเคราะห์ที่รวบรวมความสัมพันธ์ที่ไม่เป็นเชิงเส้นในระดับสูงของ SVM และ SVR ทำให้แบบจำลองเป็นที่นิยมในการศึกษาเศรษฐศาสตร์และการเงิน (Puttanapong et al., 2020)

ในการศึกษานี้ แบบจำลอง SVR จะสามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูลใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูล โดยจะจำแนกข้อมูลออกเป็น 2 ส่วน ซึ่งจะทำการสอน (Train) ข้อมูล Google Trends ให้เกิดการรู้จำของข้อมูลซึ่งเป็นชุดข้อมูลที่ต้องการจำแนก และนำข้อมูลดังกล่าว ป้อนเข้าใน SVR เพื่อคัดแยกกลุ่มข้อมูล ทั้งนี้ โครงสร้างข้อมูลสำหรับสอนหรือป้อนเข้า และผลลัพธ์ของ SVR จะทำให้ระบบเกิดการรู้จำ ดังสมการดังนี้

$$(x_i, y_i), \ldots, (x_n, y_n) \quad \text{เมื่อ} \quad x \in R^m, y \in \{+1, -1\}$$

โดยที่ $(x_i, y_i), \ldots, (x_n, y_n)$ คือ คุณลักษณะเด่นสำหรับใช้ในการสอนข้อมูล

n คือ จำนวนข้อมูลตัวอย่าง

m คือ มิติข้อมูลขาเข้า

Ref. code: 25646304010116LCN

y คือ ผลลัพธ์ที่มีค่า +1 ถึง -1

จากนั้นข้อมูลจะถูกจำแนกออกเป็นสองกลุ่ม ดังนี้

$$(w \cdot x) + b > 0 \quad \text{ถ้า} \quad y_i = +1$$
$$\text{และ } (w \cdot x) + b < 0 \quad \text{ถ้า} \quad y_i = -1$$

โดยที่ w คือ ค่าถ่วงน้ำหนัก

b คือ ค่า bias ที่มีเส้นแบ่งหรือระนาบตัดสินใจที่คำนวณได้

$$(w \cdot x) + b = 0$$

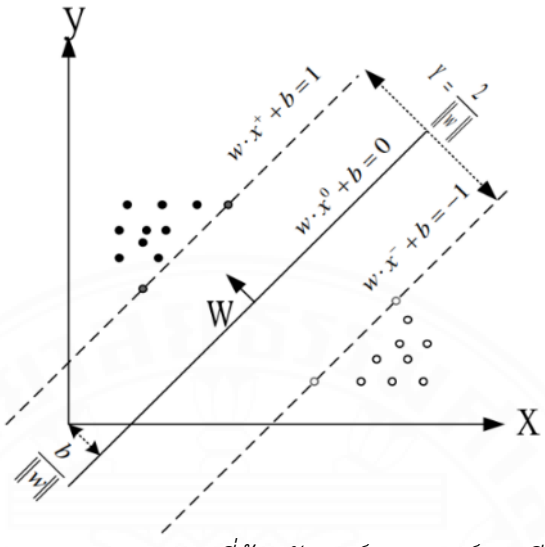
เมื่อเวกเตอร์ของข้อมูลที่ถูกนำเข้าสู่ระบบการ Train เพื่อให้ระบบเรียนรู้แทนด้วยสมการ และข้อมูลทั้ง 2 ด้านแบ่งเป็นบวกหรือลบ จากนั้นจะหาเส้นแบ่งที่ดีที่สุดของข้อมูลทั้ง 2 ด้านนั้น โดยการสร้างขอบให้กับเส้นแบ่งทั้ง 2 ข้าง ทำให้ได้เส้นขอบ (Margin) ใหม่ ซึ่งจะเป็นขอบของข้อมูลแต่ละฝั่ง และเส้นขอบที่เป็นเส้นแบ่งจะเป็นเส้นที่สัมผัสกับค่าข้อมูลใน Feature space ที่ใกล้ที่สุด และเส้นขอบทั้ง 2 เส้น จะถูกแทนด้วยสมการ ดังนี้

$$\text{ถ้าอยู่ด้าน } y = 1 : w \cdot x^{+} + b \geq y \geq 1$$
$$\text{ถ้าอยู่ด้าน } y = -1 : w \cdot x^{-} + b \leq y \leq -1$$

Ref. code: 25646304010116LCN

ภาพที่ 3.5

Support vector machine



หมายเหตุ. จาก *การคัดแยกความสุกสตรอเบอรี่ด้วยซัพพอร์ตเวกเตอร์แมชชีน*, โดย เทอดศักดิ์ เงินมูล, พิเชษฐ เหมยคำ, วิโรจน์ ปงลังกา และคณะ, 2560, Naresuan University Engineering Journal, 12(2), 58-59.

ทั้งนี้ หลักการของ SVR จะต้องการหาเส้นแบ่งที่มีขอบกว้างที่สุด เพื่อให้ของมูลทั้ง 2 ชุด มีการแบ่งแยกกันอย่างชัดเจน ดังนั้น เส้นแบ่งที่มีขอบกว้างที่สุดจะเป็นเส้นแบ่งที่ดีที่สุด (Maximization of Margin) ซึ่งสามารถหาได้จากสมการ ดังนี้

$$Y = \frac{2}{\|w\|}$$

$$w = \sum_{i=1}^{N} \alpha_i \, y_i x_i$$

โดยที่ α คือ สัมประสิทธิ์คงที่ $\alpha_i \geq 0; \;\; i = 1,2,3, \ldots, N$

$$b = -\frac{\max_{y_i=-1}(w \cdot x_i) + \min_{y_i=1}(w \cdot x_i)}{2}$$

3.3.4 แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN)

แบบจำลอง Neural Network หรือโครงข่ายประสาทเทียมมีโครงสร้างเครือข่ายที่ประกอบด้วยโหนด (Nodes) และขอบ (Edges) จำนวนมาก ซึ่งโหนดจะเป็นตัวแปรหรือฟังก์ชันทางคณิตศาสตร์ในแบบจำลองนี้ กล่าวคือเป็นการเชื่อมต่อโครงสร้างของโหนด รูปแบบของการเชื่อมต่อ (Connectivity) เช่น รูปแบบของขอบ โดยปกติแล้วโหนดขาเข้า (Input nodes) จะได้รับข้อมูลขา

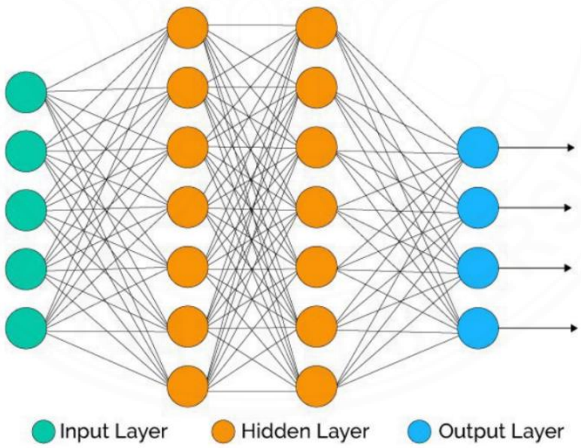
Ref. code: 25646304010116LCN

เข้า (Input data) เช่น ชั้นขาเข้า (Input layer) จากนั้น ขั้นตอนการคำนวณต่อไปจะเกิดขึ้นในชั้น (layer) ที่สอง โดยที่ข้อมูลขาเข้าจะได้รับการประมวลผลทางคณิตศาสตร์ผ่านโหนดและขอบในชั้นซ่อนเร่น (Hidden layer) อันแรก การคำนวณสามารถทำซ้ำได้ในชั้นซ่อนเร่นในภายหลัง จากนั้นชั้นสุดท้ายจะรวมค่าทั้งหมดเพื่อกำหนดการพยากรณ์ขั้นสุดท้าย (Puttanapong et al., 2020)

โดยในการศึกษานี้จะใช้แบบจำลองทางคณิตศาสตร์ที่ใช้หลักการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) เพื่อจำลองลักษณะการทำงานให้คล้ายกับโครงข่ายประสาทของมนุษย์ ประกอบด้วย 3 ส่วนสำคัญ คือ เซลล์ประสาท (Neuron หรือ Soma) ใยประสาท (Never fiber) และไซแนปส์ (Synapse) เพื่อใช้สร้างเครื่องมือที่จะเรียนรู้หรือจดจำรูปแบบ โดยหลักการทำงานแบ่งออกได้เป็น 3 ชั้น ได้แก่ ตัวแปรขาเข้าในชั้นขาเขา (Input layer) จำนวนหนึ่งชั้น สำหรับนำเข้าข้อมูลเพื่อเรียนรู้, ชั้นซ่อนเร่น (Hidden layer) เป็นชั้นตรงกลางโดยภายในจะประกอบไปด้วยโหนดสำหรับใช้ในการประมวลผล และตัวแปรขาออกในชั้นขาออก (Output layer) จำนวนหนึ่งชั้น ซึ่งได้จากการประมวลผลในชั้นแอบแฝง (ผุสดี บุญรอด, 2560) ดังภาพที่ 3.6

ภาพที่ 3.6

โครงข่ายประสาทเทียมแบบหลายชั้น



หมายเหตุ. โดย Mc, 2017, (https://towardsdatascience.com/machine-learning-fundamentals-ii-neural-networks-f1e7b2cb3eef)

โดยในงานศึกษานี้ผู้ศึกษาได้กำหนดตัวแปรและจำนวนของตัวแปรในแบบจำลอง Neural Network ไว้ ดังนี้

Ref. code: 25646304010116LCN

3.3.4.1 จำนวนตัวแปรขาเข้า (input)

จำนวนตัวแปรขาเข้าจะใช้คำค้นหาที่เกี่ยวข้องกับการหางานทั้งหมดจากข้อมูลของ Google Trends จำนวน 24 คำ

3.2.2.2 จำนวนชั้นซ่อนเร้น (hidden layer)

จำนวนชั้นซ่อนเร้นจะถูกกำหนดหาค่าที่เหมาะสมจากการคำนวณเพื่อให้ได้ผลลัพธ์ที่มีความคาดเคลื่อนน้อยลง

3.2.2.3 จำนวนตัวแปรขาออก (output)

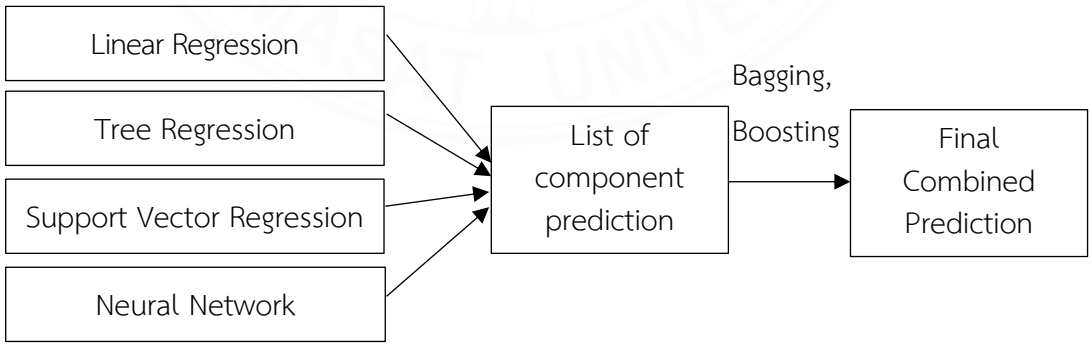
จำนวนตัวแปรขาออกจะกำหนดให้เป็น อัตราการว่างงานจากการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (NSO) หรืออัตราการว่างงานจากผู้ประกันตนในประกันสังคม (SSO)

3.3.5 แบบจำลองแบบเรียงซ้อน (Ensemble Model: EM)

เป็นวิธีการที่อาศัยแบบจำลองการจำแนกข้อมูลมากกว่าหนึ่งตัว (Base Classification) ซึ่งแต่ละแบบจำลองการจำแนกข้อมูลจะมีกระบวนการทำงานด้วยตัวเอง และทุกตัวของแบบจำลองการจำแนกจะสร้างจากกลุ่มข้อมูลเดียวกันเมื่อได้ผลลัพธ์ของแต่ละแบบจำลอง จะนำผลลัพธ์ที่ได้เหล่านั้นผ่านวิธีการรวบรวม (Combination, Integration หรือ Vote) และ ตัดสินผลลัพธ์สุดท้าย (Final Decision) เพื่อให้ได้ผลลัพธ์การจำแนกข้อมูลเดียวเท่านั้น (ปรเมษฐ์ ธันวานนท์ และคณะ, 2560)

ภาพที่ 3.7

ขั้นตอนการรวมแบบจำลองการพยากรณ์ของ Ensemble



หมายเหตุ. รวบรวมโดยผู้ศึกษา

โดยในการศึกษานี้จะเป็นการใช้ Machine Learning ที่นำแบบจำลองหลาย ๆ แบบจำลองมารวมกัน โดยจะใช้อัลอกริทึมของ Machine Learning เช่น Linear Regression

Ref. code: 25646304010116LCN

Model, Optimizable Tree Regression Model, Optimizable Support Vector Regression Model, Neural Network Model และ อื่น ๆ ดังภาพที่ 3.7 เพื่อที่จะทำให้ผลลัพธ์หรือประสิทธิภาพออกมาดีที่สุด ซึ่งในงานศึกษานี้จะใช้เครื่องมือ 2 แบบ ได้แก่ Bagging คือการสร้างหลายชุดข้อมูลแต่ประเภทเดียวกัน เพื่อทำการทดสอบกับแต่ละชุดข้อมูลย่อยที่แบ่งออกจากชุดข้อมูลรวม จากนั้นนำผลการพยากรณ์ของแบบจำลองต่าง ๆ มารวมกัน และ Boosting ที่ใช้รูปแบบการสุ่มเลือกข้อมูลที่เรียกว่า Bootstrapping คือ การสุ่มเลือกข้อมูลที่มีขั้นตอนโดยการสุ่มเลือกข้อมูลจากชุดข้อมูลทั้งหมด ซึ่งข้อมูลที่เคยถูกสุ่มเลือกออกมาแล้ว ข้อมูลนั้นจะถูกเก็บไว้ในชุดข้อมูลใหม่ หลังจากนั้นจะทำการนำข้อมูลที่ถูกสุ่มเลือกเก็บกลับไปยังชุดข้อมูลทั้งหมดตามเดิม ทำให้ข้อมูลที่เคยถูกสุ่มเลือกมีโอกาสที่จะถูกสุ่มเลือกใหม่อีกครั้งในการสุ่มเลือกครั้งถัดไป โดยขั้นตอนการสุ่มเลือกข้อมูลแบบ Bagging หรือ Boosting เมื่อสุ่มจนได้จำนวนชุดข้อมูลตามที่ต้องการเรียบร้อยแล้ว จะทำการนำชุดข้อมูลที่ได้ทั้งหมดไปทำการเรียนรู้เพื่อสร้างแบบจำลองการพยากรณ์ โดยการใช้อัลกอริทึมการสร้างแบบจำลองการพยากรณ์เพียงอัลกอริทึมเดียวเท่านั้น ซึ่งในขั้นตอนสุดท้ายจะเป็นการรวบรวมคำตอบในแต่ละแบบจำลองมาพยากรณ์

3.4 การประเมินประสิทธิภาพของแบบจำลอง

โดยแบบจำลองที่กล่าวมาข้างต้น จะนำมาตรวจสอบความแม่นยำของแบบจำลอง เพื่อหาแบบจำลองที่เหมาะสมที่สุดในการพยากรณ์อัตราการว่างงานจากข้อมูล Google Trends จะมีรายละเอียดขั้นตอนในการศึกษา ดังนี้

เป็นการนำการพยากรณ์ที่ได้จากแบบจำลองทั้ง 5 วิธีมาเปรียบเทียบกับค่าจริงที่เกิดขึ้น ซึ่งใช้ชุดข้อมูล Testing Data ที่คิดเป็นร้อยละ 20 ของข้อมูลทั้งหมด แล้วคำนวณโดยใช้วิธี Root Mean Square Error (RMSE) เพื่อเป็นการหาความผิดพลาดของแบบจำลอง โดยมีสมการ ดังนี้

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \widehat{y_t})^2}{n}}$$

โดยที่ y_t คือ ข้อมูลจริง ณ เวลาที่ t

$\widehat{y_t}$ คือ ค่าพยากรณ์ของข้อมูล ณ เวลาที่ t

n คือ จำนวนข้อมูล

Ref. code: 25646304010116LCN

สำหรับหลักเกณฑ์การตัดสินใจเลือกแบบจำลองจะเลือกแบบจำลองที่มีค่า RMSE น้อยที่สุด ซึ่งแสดงว่าแบบจำลองนั้นมีค่าการพยากรณ์ที่คาดเคลื่อนไปจากค่าจริงน้อยที่สุด

Ref. code: 25646304010116LCN

บทที่ 4
ผลการวิจัยและอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาหาความสัมพันธ์ระหว่างคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends กับอัตราการว่างงานในประเทศไทย พร้อมทั้งใช้ข้อมูลคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends นำไปพยากรณ์อัตราการว่างงานในประเทศไทย และสามารถหาแบบจำลองที่มีประสิทธิภาพที่สุดด้วยวิธีการพยากรณ์ที่แตกต่างกันออกไป โดยจะใช้แบบจำลอง 5 วิธี ได้แก่ Linear Regression Model, Optimizable Tree Regression Model, Optimizable Support Vector Regression Model, Neural Network Model และ Optimizable Ensemble Model ทั้งนี้ จะใช้อัลกอริทึมของ Machine Learning ในรูปของ Regression Learner เพื่อหาแบบจำลองที่มีประสิทธิภาพที่สุด ส่วนข้อมูลอัตราการว่างงานมาจาก 2 หน่วยงานภาครัฐ ส่วนแรกเป็นอัตราการว่างงานเก็บข้อมูลจากการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (NSO) รายเดือน (ปี พ.ศ. 2554 - 2563 โดยในปี พ.ศ. 2563 โดยช่วงเดือนเมษายน - มิถุนายน จะไม่มีการเก็บข้อมูล) อีกส่วนเป็นอัตราการว่างงานเก็บข้อมูลจากผู้ประกันตนในประกันสังคม (SSO) ที่ขอรับเงินชดเชยกรณีว่างงานรายเดือน (เดือนมกราคม พ.ศ. 2561 – ตุลาคม พ.ศ. 2564) จากนั้นนำผลของการพยากรณ์ในแต่ละแบบจำลองมาเปรียบเทียบกับค่าที่เกิดขึ้นจริงแล้วคำนวณหาค่า RMSE (Root Mean Square Error) เพื่อหาแบบจำลองที่ให้ความแม่นยำมากที่สุด โดยในงานวิจัยนี้ได้แบ่งการศึกษาออกเป็น 2 ส่วน ดังนี้

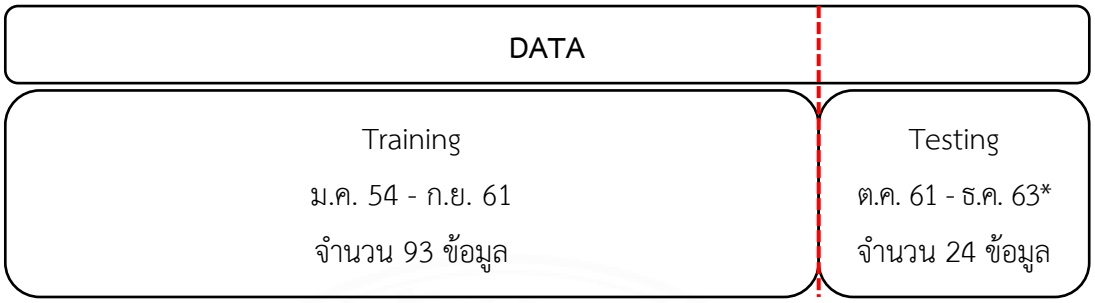
4.1 ผลลัพธ์ของแบบจำลองอัตราการว่างงาน

นำข้อมูลจาก Google Trends และอัตราการว่างงานจากการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (NSO) และผู้ประกันตนในประกันสังคม (SSO) สร้างแบบจำลองเพื่อพยากรณ์อัตราการว่างงาน โดยใช้โปรแกรม MATLAB ซึ่งจะแบ่งข้อมูลตัวแปรต้น หรือ Google Trends ที่จะศึกษาออกเป็น 2 ชุดข้อมูล โดยวิธี Split Test ในสัดส่วน 80 : 20 กล่าวคือ ข้อมูลที่ถูกนำไปใช้ในการเรียนรู้เพื่อสร้างแบบจำลอง (Training Data) คิดเป็นร้อยละ 80 และข้อมูลที่ใช้ทดสอบความถูกต้องของแบบจำลอง (Testing Data) คิดเป็นร้อยละ 20 โดยจะเรียงข้อมูลอัตราการว่างงานและ Google Trends ตามรายเดือนในช่วงเวลาเดียวกัน มีรายละเอียดในการแบ่งข้อมูล ดังนี้

Ref. code: 25646304010116LCN

ภาพที่ 4.1

การแบ่งข้อมูลของชุดข้อมูลการสำรวจประชากรไทยโดยสำนักงานสถิติแห่งชาติ (NSO)

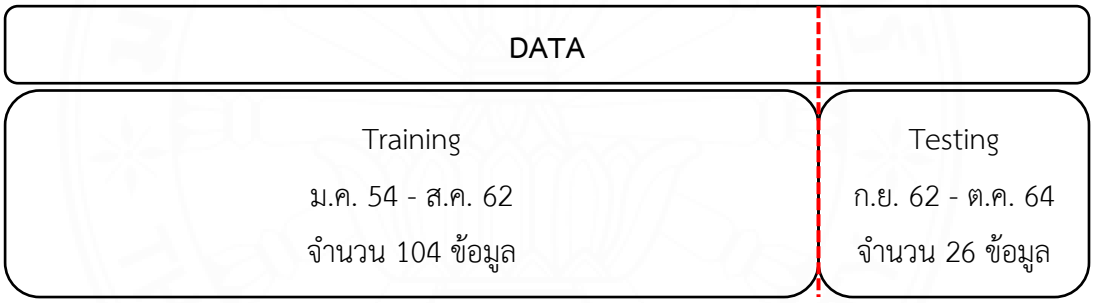


* ช่วงเดือนเมษายน – มิถุนายน 2563 จะไม่มีการเก็บข้อมูล

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.2

การแบ่งข้อมูลของชุดข้อมูลผู้ประกันตนในประกันสังคม (SSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

และกำหนดตัวแปร Google Trends ที่ใช้ศึกษาต่าง ๆ เพื่อใช้ในโปรแกรม MATLAB แสดงรายละเอียดดังตารางที่ 4.1 ดังนี้

ตารางที่ 4.1

กำหนดตัวแปร Google Trends ที่ใช้ในการศึกษา

คำค้นหา Google Trends	ตัวแปร	ชื่อตัวแปร
สมัครงานบริษัท	ApplyForACompanyJob	X1
หางาน	ApplyForWork	X2
ว่างงาน	Available	X3
เงินชดเชยเลิกจ้าง	Compensation	X4
กรมการจัดหางาน	DOE	X5

Ref. code: 25646304010116LCN

ตารางที่ 4.1

กำหนดตัวแปร Google Trends ที่ใช้ในการศึกษา (ต่อ)

คำค้นหา Google Trends	ตัวแปร	ชื่อตัวแปร
นายจ้างหาคนงาน	EmployersLookingForWorkers	X6
สมัครงาน	FindAJob	X7
หางานสมัครงาน	FindAJobApplyForAJob	X8
หางานทำ	FindWork	X9
งานราชการ	GovernmentJob	X10
สัมภาษณ์งาน	Interview	X11
ใบสมัครงาน	JobApplication	X12
jobfairไทย	Jobfair	X13
jobsdbไทย	Jobsdb	X14
jobthไทย	JobTH	X15
jobthaiไทย	JobThai	X16
jobtopgunไทย	JobTopGun	X17
linkedinไทย	LinkedIn	X18
กองทนสำรองเลี้ยงชีพ	ProvidentFund	X19
ประกันสังคม	SocialSecurity	X20
สิทธิประกันสังคม	SocialSecurityRights	X21
ตกงาน	Unemploy	X22
เงินชดเชยว่างงาน	UnemploymentCompensation	X23
ลงทะเบียนว่างงาน	UnemploymentRegistration	X24

หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากนั้นนำชุดข้อมูลที่ใช้ในการเรียนรู้เพื่อสร้างแบบจำลอง (Training Data) ไปทำการสร้างสมการแบบจำลองวิธีต่าง ๆ และนำข้อมูลที่ใช้ทดสอบความถูกต้องของแบบจำลอง (Testing Data) มาทดสอบในสมการดังกล่าว เพื่อทดสอบค่าความถูกต้องต่อไป ซึ่งผลลัพธ์ของแต่ละแบบจำลองสามารถอธิบายได้ ดังนี้

Ref. code: 25646304010116LCN

4.1.1 แบบจำลองการถดถอยเชิงเส้น (Linear Regression Model)

เมื่อนำชุดข้อมูล Training Data ไปสร้างแบบจำลองและใช้ชุดข้อมูล Testing Data เพื่อทดสอบแบบจำลอง โดยในการศึกษานี้จะใช้เงื่อนไขของ Linear Regression ซึ่งจะพิจารณาความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามที่เป็นความสัมพันธ์เชิงเส้นตรง ซึ่งจะใช้โปรแกรม MATLAB ในการสร้างแบบจำลองที่ใช้ชื่อว่า Linear Model ได้ผลลัพธ์ในแต่ละชุดข้อมูลดังนี้

4.1.1.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)

สร้างแบบจำลองการถดถอยเชิงเส้นจากการนำชุดข้อมูล Training Data ทำให้ได้แบบจำลองของข้อมูลชุดเรียนรู้ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยโปรแกรม MATLAB โดยจะใช้อัลกอริทึมของ Machine Learning ในรูปของ Regression Learner เพื่อให้ได้แบบจำลองที่เหมาะสม ทั้งนี้ ผลลัพธ์ของการสร้างแบบจำลองดังกล่าวสามารถแสดงรายละเอียดผลการศึกษาปัจจัยที่มีอิทธิพลของ Google Trends ต่ออัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ในรูปของค่าสถิติเบื้องต้นได้ ดังนี้

ตารางที่ 4.2

ค่าสถิติเบื้องต้นของตัวแปรที่ศึกษา (Descriptive Statistics) ในชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธี Linear Regression

Unemployment rate	Coef.	Std. Err.	t	P>|t|
x1	-0.026701	0.028546	-0.935360	0.352910
x2	0.026436	0.024551	1.076800	0.285390
x3	-0.082003	0.084569	-0.969660	0.335650
x4	0.006092	0.019740	0.308600	0.758570
x5	-0.013223	0.009765	-1.354100	0.180180
x6	0.030341	0.093272	0.325290	0.745960
x7	0.005489	0.017452	0.314570	0.754050
x8	0.080380	0.042586	1.887500	0.063364
x9	-0.031374	0.050441	-0.622010	0.536020
x10	-0.005347	0.003564	-1.500200	0.138180
x11	-0.033779	0.032485	-1.039900	0.302090
x12	-0.097797	0.065373	-1.496000	0.139290

Ref. code: 25646304010116LCN

ตารางที่ 4.2

ค่าสถิติเบื้องต้นของตัวแปรที่ศึกษา (Descriptive Statistics) ในชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธี Linear Regression (ต่อ)

Unemployment rate	Coef.	Std. Err.	t	P>|t|
x13	0.005669	0.018693	0.303290	0.762590
X14	-0.013970	0.044200	-0.316060	0.752920
x15	0.025626	0.021510	1.191300	0.237660
x16	0.011341	0.009789	1.158600	0.250680
x17	-0.033801	0.031123	-1.086000	0.281290
x18	0.006915	0.023212	0.297920	0.766670
x19	-0.002833	0.007022	-0.403450	0.687890
x20	0.026936	0.015095	1.784500	0.078809
x21	-0.047045	0.064399	-0.730520	0.467580
x22	0.094349	0.084027	1.122800	0.265450
x23	0.077479	0.053737	1.441800	0.153940
x24	0.029088	0.012305	2.363800	0.020953
(Intercept)	0.387810	0.156800	2.473400	0.015888

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ทั้งนี้ จากตารางที่ 4.2 แสดงผลการศึกษาแบบจำลองการพยากรณ์อัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธี Linear Regression สามารถสรุปรายละเอียดของแบบจำลอง ดังตารางที่ 4.3

Ref. code: 25646304010116LCN

ตารางที่ 4.3

รายละเอียดแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)

Source	SS	DF	MS
Model	3.481400	24	0.145060
Residual	1.280600	68	0.018832
Total	**4.762000**	**92**	**0.051761**

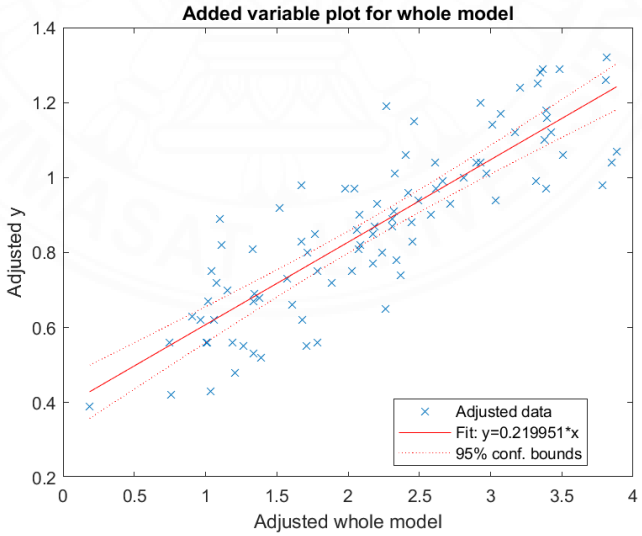
Number of obs	=	93
F (24, 68)	=	7.702900
Prob > F	=	1.4776E-11
R-squared	=	0.731000
Adj R-squared	=	0.636000
Root MSE	=	0.137000

หมายเหตุ. รวบรวมโดยผู้ศึกษา

นอกจากนี้ ยังสามารถแสดงการเปรียบเทียบระหว่างค่าพารามิเตอร์หรือตัวแปร Google Trends กับเส้นสมการเชิงเส้นตรงและมีช่วงความเชื่อมั่น (Confidence Interval) ที่ร้อยละ 95 ดังภาพที่ 4.3

ภาพที่ 4.3

ค่าพารามิเตอร์กับแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานจากสำนักงานสถิติแห่งชาติ (NSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากการวิเคราะห์แบบจำลองการพยากรณ์อัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Linear Regression พบว่า ตัวแปร Google Trends มีผล

Ref. code: 25646304010116LCN

การทดสอบด้วยค่าสถิติ P-Value ที่ระดับนัยสำคัญทางสถิติที่ 0.1 มีเพียงคำว่า “หางานสมัครงาน”, “ประกันสังคม” และ“ลงทะเบียนว่างงาน” กล่าวคือ ณ ระดับนัยสำคัญทางสถิติที่ 0.1 ผลลัพธ์ของค่าสัมประสิทธิ์ที่คำนวณจากข้อมูล Training Data สามารถใช้แทนสัมประสิทธิ์ในสมการถดถอยของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ได้

อย่างไรก็ตาม เมื่อนำตัวแปร Google Trends ของข้อมูลทั้ง 24 ตัวแปร มาสร้างแบบจำลองการพยากรณ์อัตราการว่างงานด้วยวิธีการ Linear Regression มาทดสอบด้วยค่าสถิติ F-Stat พบว่า มีค่า P-Value เท่ากับ 1.4776E-11 กล่าวคือ ณ ระดับนัยสำคัญทางสถิติที่ต่ำกว่าระดับ 0.0001 ผลลัพธ์ของค่าสัมประสิทธิ์ของข้อมูลทั้ง 24 ตัวแปร ที่คำนวณจากข้อมูล Training Data สามารถใช้แทนสัมประสิทธิ์ในสมการถดถอยของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ได้

4.1.1.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

สร้างแบบจำลองการถดถอยเชิงเส้นจากการนำชุดข้อมูล Training Data ทำให้ได้แบบจำลองของข้อมูลชุดเรียนรู้ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยโปรแกรม MATLAB โดยจะใช้อัลกอริทึมของ Machine Learning ในรูปของ Regression Learner เพื่อให้ได้แบบจำลองที่เหมาะสม ทั้งนี้ ผลลัพธ์ของการสร้างแบบจำลองดังกล่าวสามารถแสดงรายละเอียดผลการศึกษาปัจจัยที่มีอิทธิพลของ Google Trends ต่ออัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ในรูปของค่าสถิติเบื้องต้นได้ ดังนี้

ตารางที่ 4.4

สรุปค่าสถิติเบื้องต้นของตัวแปรที่ศึกษา (Descriptive Statistics) ในชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธี Linear Regression

Unemployment rate	Coef.	Std. Err.	t	P>|t|
x1	-0.027599	0.020946	-1.317700	0.191420
x2	0.003041	0.018427	0.165010	0.869360
x3	0.040729	0.052852	0.770640	0.443220
x4	-0.016133	0.013533	-1.192100	0.236780
x5	0.004692	0.006534	0.718070	0.474830
x6	0.017050	0.066367	0.256900	0.797920
x7	-0.012790	0.012466	-1.026100	0.308000
x8	0.054316	0.030836	1.761500	0.082030

Ref. code: 25646304010116LCN

ตารางที่ 4.4

สรุปค่าสถิติเบื้องต้นของตัวแปรที่ศึกษา (Descriptive Statistics) ในชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธี Linear Regression (ต่อ)

Unemployment rate	Coef.	Std. Err.	t	P>|t|
x9	0.046041	0.035555	1.294900	0.199120
x10	0.004840	0.002641	1.832900	0.070587
x11	0.002613	0.024005	0.108850	0.913600
x12	0.000508	0.049694	0.010218	0.991870
x13	-0.012677	0.014175	-0.894330	0.373860
X14	-0.078354	0.033108	-2.366700	0.020399
x15	0.018084	0.016370	1.104700	0.272630
x16	0.000614	0.007237	0.084860	0.932590
x17	-0.063080	0.022697	-2.779200	0.006806
x18	0.003092	0.016902	0.182940	0.855310
x19	0.005813	0.005038	1.153700	0.252100
x20	0.002425	0.010956	0.221350	0.825390
x21	-0.018833	0.045328	-0.415490	0.678910
x22	0.024995	0.063949	0.390860	0.696950
x23	0.008138	0.037818	0.215180	0.830180
x24	0.014541	0.008047	1.807100	0.074554
(Intercept)	0.810860	0.113380	7.151600	3.828E-10

หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากตารางที่ 4.4 แสดงผลการศึกษาแบบจำลองการพยากรณ์อัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธี Linear Regression สามารถสรุปรายละเอียดของแบบจำลอง ดังตารางที่ 4.5

Ref. code: 25646304010116LCN

ตารางที่ 4.5

รายละเอียดแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

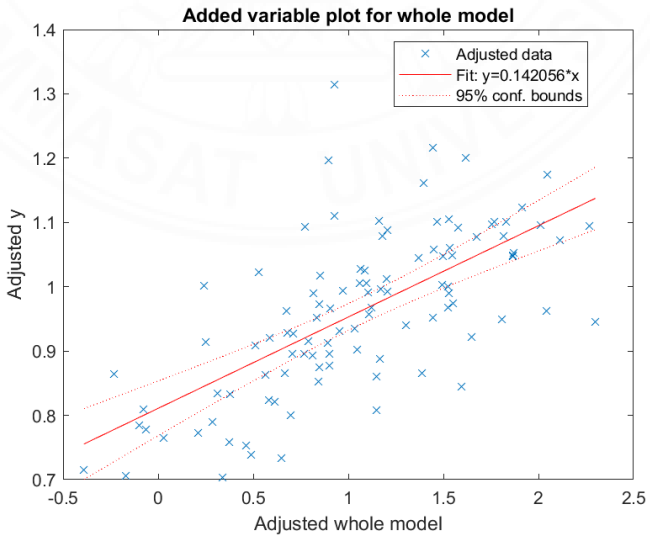
Source	SS	DF	MS			
Model	0.723610	24	0.030151	Number of obs	=	104
Residual	0.891940	79	0.011290	F (24, 79)	=	2.670500
Total	**1.615600**	**103**	**0.015685**	Prob > F	=	0.000581
				R-squared	=	0.448000
				Adj R-squared	=	0.280000
				Root MSE	=	0.106000

หมายเหตุ. รวบรวมโดยผู้ศึกษา

นอกจากนี้ ยังสามารถแสดงการเปรียบเทียบระหว่างค่าพารามิเตอร์หรือตัวแปร Google Trends กับเส้นสมการเชิงเส้นตรงและมีช่วงความเชื่อมั่น (confidence interval) ที่ร้อยละ 95 ดังภาพที่ 4.4

ภาพที่ 4.4

ค่าพารามิเตอร์กับแบบจำลองการถดถอยเชิงเส้นของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO).



หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

จากการวิเคราะห์แบบจำลองการพยากรณ์อัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Linear Regression พบว่า ตัวแปร Google Trends มีผลการทดสอบด้วยค่าสถิติ P-Value ที่ระดับนัยสำคัญทางสถิติที่ 0.1 มีเพียงคำว่า “หางานสมัครงาน”, “งานราชการ”, “JobDB”, “JobTopGun” และ “ลงทะเบียนว่างงาน” กล่าวคือ ณ ระดับนัยสำคัญทางสถิติที่ 0.1 ผลลัพธ์ของค่าสัมประสิทธิ์ที่คำนวณจากข้อมูล Training Data สามารถใช้แทนสัมประสิทธิ์ในสมการถดถอยของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ได้

อย่างไรก็ตาม เมื่อนำตัวแปร Google Trends ของข้อมูลทั้ง 24 ตัวแปร มาสร้างแบบจำลองการพยากรณ์อัตราการว่างงานด้วยวิธีการ Linear Regression มาทดสอบด้วยค่าสถิติ F-Stat พบว่า มีค่า P-Value เท่ากับ 5.8147E-04 กล่าวคือ ณ ระดับนัยสำคัญทางสถิติที่ต่ำกว่า 0.001 ผลลัพธ์ของค่าสัมประสิทธิ์ของข้อมูลทั้ง 24 ตัวแปร ที่คำนวณจากข้อมูล Training Data สามารถใช้แทนสัมประสิทธิ์ในสมการถดถอยของผู้ประกันตนในประกันสังคม (SSO) ได้

4.1.2 แบบจำลองต้นไม้ตัดสินใจ (Optimizable Decision Tree Regression Model)

เมื่อนำชุดข้อมูล Training Data ไปสร้างแบบจำลองและใช้ชุดข้อมูล Testing Data เพื่อทดสอบแบบจำลอง โดยในการศึกษานี้จะใช้เงื่อนไขของ Optimizable Tree Regression ซึ่งจะทำการพยากรณ์อัตราการว่างงานโดยใช้ค่าเฉลี่ยของค่าอัตราการว่างงาน และจะทำการแบ่งข้อมูลจากโหนดรากไปจนถึงโหนดใบ จนกว่าจะได้ตามเงื่อนไขที่กำหนด ซึ่งจะใช้โปรแกรม MATLAB ในการสร้างแบบจำลองที่ใช้ชื่อว่า Optimizable Tree Model ได้ผลลัพธ์ในแต่ละชุดข้อมูล ดังนี้

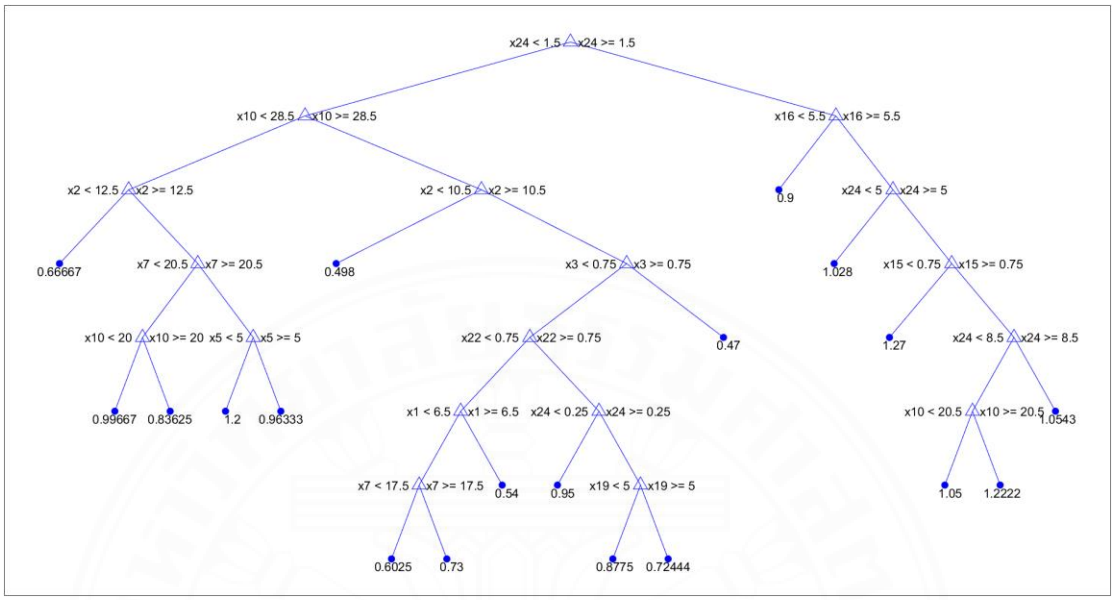
4.1.2.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)

จากการนำกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) มาทดสอบกับอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) ทำให้ได้แบบจำลองของข้อมูลชุดเรียนรู้ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยโปรแกรม MATLAB เพื่อให้ได้แบบจำลองที่เหมาะสม ทั้งนี้ ผลลัพธ์ของการพยากรณ์อัตราการว่างงานจากแบบจำลองต้นไม้ตัดสินใจที่ใช้ข้อมูลจากชุดข้อมูล Google Trends ในช่วงเวลาเดียวกันกับอัตราการว่างงานในแต่ละเดือน สามารถแสดงรายละเอียดของแบบจำลองได้ตามภาพที่ 4.5 ดังนี้

Ref. code: 25646304010116LCN

ภาพที่ 4.5

แผนภูมิต้นไม้ตัดสินใจของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO).



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากการวิเคราะห์แบบจำลองการพยากรณ์อัตราการว่างงานของอัตราการว่างงานสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการต้นไม้ตัดสินใจ พบว่า เงื่อนไขแรกที่เป็นจุดเริ่มต้นคือ ตัวแปร Google Trends คำว่า “ลงทะเบียนว่างงาน” เมื่อมีค่าน้อยกว่า 1.5 จะแตกกิ่งไปสู่เงื่อนไขถัดไป คือคำว่า “งานราชการ” หรือเมื่อมีค่ามากกว่าเท่ากับ 1.5 จะแตกกิ่งไปสู่เงื่อนไขถัดไป คือคำว่า “JobThai” และจะแตกกิ่งไปสู่เงื่อนไขถัดไปเรื่อย ๆ ตามภาพที่ 4.5 จนกระทั่งครบทุกเงื่อนไขที่เกี่ยวข้องกับการตัดสินใจ ซึ่งพบว่าผลลัพธ์ของต้นไม้ตัดสินใจประกอบด้วยโหนดใบ ได้แก่

(1) คำว่า “หางาน” ถ้าน้อยกว่า 12.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.66667

(2) คำว่า “งานราชการ” ถ้าน้อยกว่า 20 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.99667 ขณะที่ถ้ามากกว่าหรือเท่ากับ 20 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.83625

(3) คำว่า “กรมการจัดหางาน” ถ้าน้อยกว่า 5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.2 ขณะที่ถ้ามากกว่าหรือเท่ากับ 5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.96333

(4) คำว่า “หางาน” ถ้าน้อยกว่า 10.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.498

Ref. code: 25646304010116LCN

(5) คำว่า “สมัครงาน” ถ้าน้อยกว่า 17.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.6025 ขณะที่ถ้ามากกว่าหรือเท่ากับ 17.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.73

(6) คำว่า “สมัครงานบริษัท” ถ้ามากกว่าเท่ากับ 6.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.54

(7) คำว่า “ลงทะเบียนว่างงาน” ถ้าน้อยกว่า 0.25 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.95

(8) คำว่า “กองทุนสำรองเลี้ยงชีพ” ถ้าน้อยกว่า 5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.8775 ขณะที่ถ้ามากกว่าหรือเท่ากับ 5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.72444

(9) คำว่า “ว่างงาน” ถ้ามากกว่าเท่ากับ 0.75 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.47

(10) คำว่า “jobthai” ถ้าน้อยกว่า 5.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.9

(11) คำว่า “ลงทะเบียนว่างงาน” ถ้าน้อยกว่า 5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.028

(12) คำว่า “jobth” ถ้าน้อยกว่า 0.75 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.27

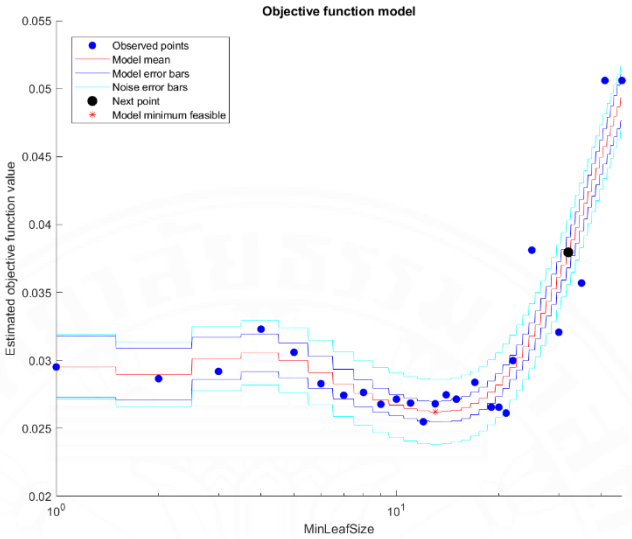
(13) คำว่า “งานราชการ” ถ้าน้อยกว่า 20.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.05 ขณะที่ถ้ามากกว่าหรือเท่ากับ 20.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.2222

(14) คำว่า “ลงทะเบียนว่างงาน” ถ้ามากกว่าเท่ากับ 8.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.0543

Ref. code: 25646304010116LCN

ภาพที่ 4.6

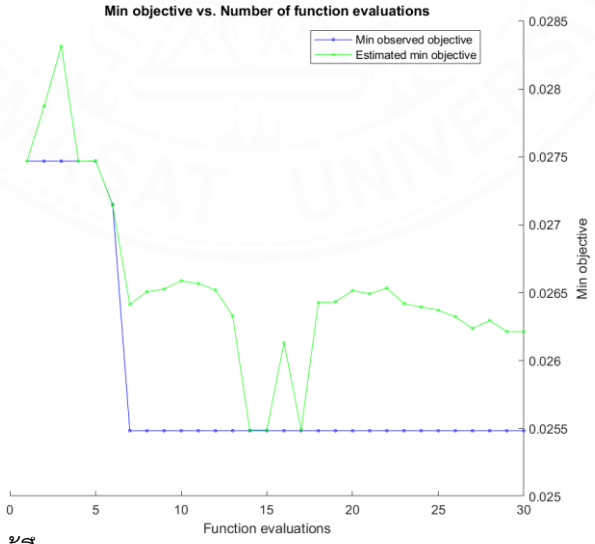
ค่าแบบจำลองของฟังก์ชันวัตถุประสงค์จากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.7

ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

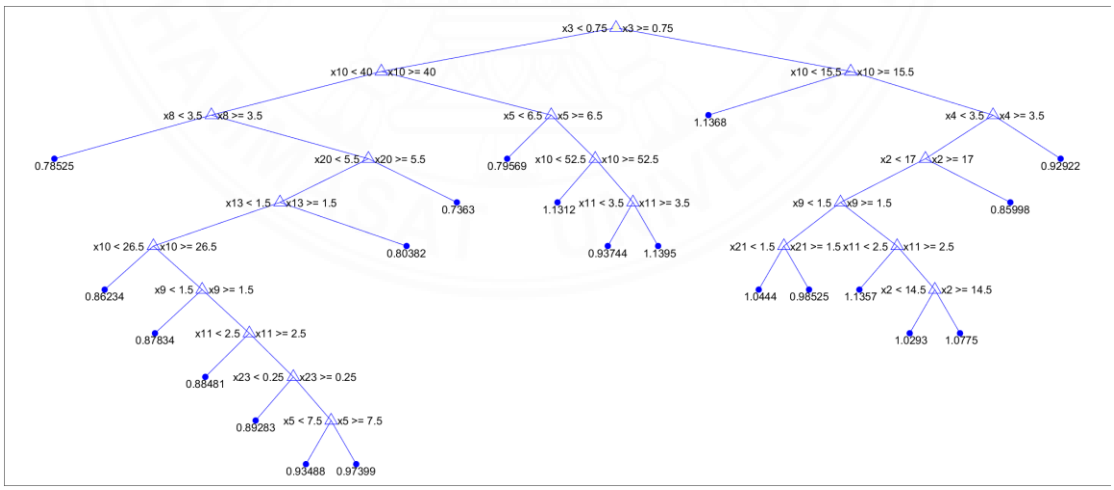
จากภาพที่ 4.6 และภาพที่ 4.7 เป็นภาพที่แสดงฟังก์ชันวัตถุประสงค์ (Objective Function) และค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ (Minimum Objective Function) ซึ่งอยู่ในรูปของการทำ Optimization และแสดงจำนวนของการประเมินค่าฟังก์ชัน (Function Evaluations) โดยจะแสดงทั้งค่าตัวอย่างที่ต่ำสุดของฟังก์ชันวัตถุประสงค์ (Minimum Observed Objective Function) และค่าพยากรณ์ที่ต่ำสุดของฟังก์ชันวัตถุประสงค์ (Estimated Observed Objective Function) จะเห็นได้ว่า ค่าพยากรณ์ต่ำที่สุดของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ค่อนข้างต่างจากเส้นต่ำสุดของฟังก์ชันวัตถุประสงค์

4.1.2.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

จากการนำกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) มาทดสอบกับอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) ทำให้ได้แบบจำลองของข้อมูลชุดเรียนรู้ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยโปรแกรม MATLAB เพื่อให้ได้แบบจำลองที่เหมาะสม ทั้งนี้ ผลลัพธ์ของการพยากรณ์อัตราการว่างงานจากแบบจำลองต้นไม้ตัดสินใจที่ใช้ข้อมูลจากชุดข้อมูล Google Trends ในช่วงเวลาเดียวกันกับอัตราการว่างงานในแต่ละเดือน สามารถแสดงรายละเอียดของแบบจำลองได้ตามภาพที่ 4.8 ดังนี้

ภาพที่ 4.8

แผนภูมิต้นไม้ตัดสินใจของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากการวิเคราะห์แบบจำลองการพยากรณ์อัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการต้นไม้ตัดสินใจ พบว่า เงื่อนไขแรกที่เป็นจุดเริ่มต้นคือ

Ref. code: 25646304010116LCN

ตัวแปร Google Trends คำว่า “ว่างงาน” เมื่อมีค่าน้อยกว่า 0.75 จะแตกกิ่งไปสู่เงื่อนไขถัดไป คือคำว่า “งานราชการ” หรือเมื่อมีค่ามากกว่าเท่ากับ 0.75 จะแตกกิ่งไปสู่เงื่อนไขถัดไป คือคำว่า “งานราชการ” เช่นกัน และจะแตกกิ่งไปสู่เงื่อนไขถัดไปเรื่อย ๆ ตามภาพที่ 4.8 จนกระทั่งครบทุกเงื่อนไขที่เกี่ยวข้องกับการตัดสินใจ ซึ่งพบว่าผลลัพธ์ของต้นไม้ตัดสินใจประกอบด้วยโหนดใบ ได้แก่

(1) คำว่า “หางานสมัครงาน” ถ้าน้อยกว่า 3.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.78525

(2) คำว่า “งานราชการ” ถ้าน้อยกว่า 26.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.86234

(3) คำว่า “หางานทำ” ถ้าน้อยกว่า 1.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.87834

(4) คำว่า “สัมภาษณ์งาน” ถ้าน้อยกว่า 2.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.88481

(5) คำว่า “เงินชดเชยว่างงาน” ถ้าน้อยกว่า 0.25 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.89283

(6) คำว่า “กรมการจัดหางาน” ถ้าน้อยกว่า 7.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.93488 ขณะที่ถ้ามากกว่าหรือเท่ากับ 7.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.97399

(7) คำว่า “jobfair” ถ้ามากกว่าเท่ากับ 1.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.80382

(8) คำว่า “ประกันสังคม” ถ้ามากกว่าเท่ากับ 5.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.7363

(9) คำว่า “กรมการจัดหางาน” ถ้าน้อยกว่า 6.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.79569

(10) คำว่า “งานราชการ” ถ้าน้อยกว่า 52.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.1312

(11) คำว่า “สัมภาษณ์งาน” ถ้าน้อยกว่า 3.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.93744 ขณะที่ถ้ามากกว่าหรือเท่ากับ 3.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.1395

(12) คำว่า “งานราชการ” ถ้าน้อยกว่า 15.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.1368

Ref. code: 25646304010116LCN

(13) คำว่า “สิทธิประกันสังคม” ถ้าน้อยกว่า 1.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.0444 ขณะที่ถ้ามากกว่าหรือเท่ากับ 1.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.98525

(14) คำว่า “สัมภาษณ์งาน” ถ้าน้อยกว่า 2.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.1357

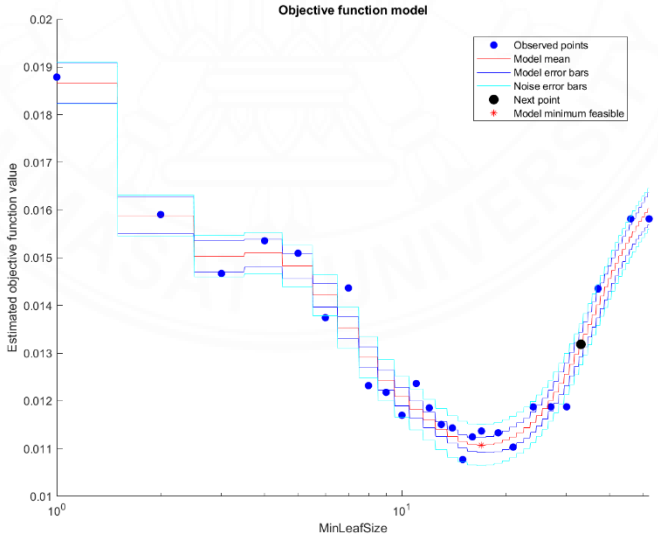
(15) คำว่า “หางาน” ถ้ามากกว่าเท่ากับ 14.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.0293 ขณะที่ถ้ามากกว่าหรือเท่ากับ 14.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 1.0775

(16) คำว่า “หางาน” ถ้ามากกว่าเท่ากับ 17 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.85998

(17) คำว่า “เงินชดเชยเลิกจ้าง” ถ้ามากกว่าเท่ากับ 3.5 จะมีผลลัพธ์ของอัตราการว่างงานเท่ากับ 0.92922

ภาพที่ 4.9

ค่าแบบจำลองของฟังก์ชันวัตถุประสงค์จากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

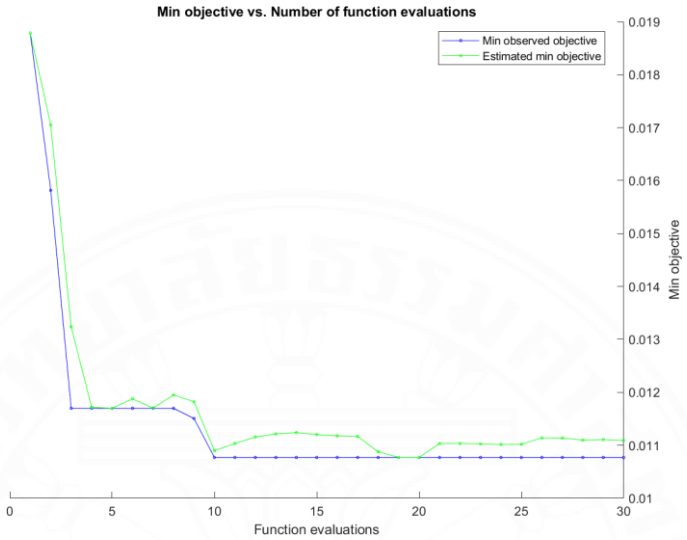


หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

ภาพที่ 4.10

ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Tree ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากภาพที่ 4.9 และ 4.10 เป็นภาพที่แสดงฟังก์ชันวัตถุประสงค์ (Objective Function) และภาพที่แสดงค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ (Minimum Objective Function) ซึ่งอยู่ในรูปของการทำ Optimization และแสดงจำนวนของการประเมินค่าฟังก์ชัน (Function Evaluations) โดยจะแสดงทั้งค่าตัวอย่างที่ต่ำสุดของฟังก์ชันวัตถุประสงค์ (Minimum Observed Objective Function) และ ค่าพยากรณ์ที่ต่ำสุดของฟังก์ชันวัตถุประสงค์ (Estimated Observed Objective Function) จะเห็นได้ว่า ค่าพยากรณ์ต่ำที่สุดของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ค่อนข้างแนบสนิทเส้นต่ำสุดของฟังก์ชันวัตถุประสงค์

4.1.3 แบบจำลองซับพอรตเวกเตอร์รีเกรสชัน (Optimizable Support Vector Regression Model: SVR)

เมื่อนำชุดข้อมูล Training Data ไปสร้างแบบจำลอง Support Vector Regression และใช้ชุดข้อมูล Testing Data เพื่อทดสอบแบบจำลอง ในงานศึกษานี้จะใช้การสอนชุดข้อมูลด้วยกระบวนการ Kernel Function โดยใช้เงื่อนไขของแบบจำลอง Support Vector Regression ซึ่งเป็นการกำหนดกลุ่มเป้าหมายเพื่อหาฟังก์ชันการตัดสินใจที่สามารถแยกค่าที่ไม่ทราบได้ โดยจะใช้โปรแกรม MATLAB ในการสร้างแบบจำลองที่ใช้ชื่อว่า Optimizable SVR Model ได้ผลลัพธ์ในแต่ละชุดข้อมูล ดังนี้

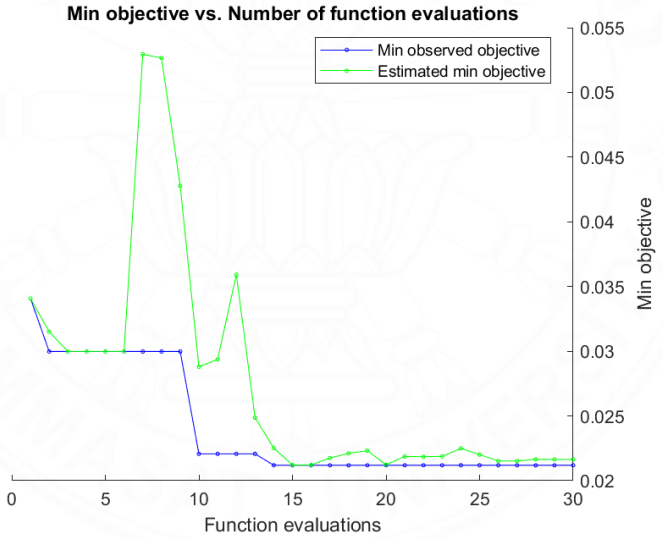
Ref. code: 25646304010116LCN

4.1.3.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)

จากการนำกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) มาผ่านกระบวนการ Support Vector Regression ทำให้ได้แบบจำลองของข้อมูลชุดเรียนรู้ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยโปรแกรม MATLAB เพื่อให้ได้แบบจำลองที่เหมาะสม และใช้ Kernel Function โดยการเพิ่ม Hyperparameter Optimization ซึ่งในการศึกษานี้จะใช้ Machine Learning ในการกำหนด ทั้งนี้ การพยากรณ์อัตราการว่างงานจากแบบจำลอง Support Vector Regression จะใช้ข้อมูลจากชุดข้อมูล Google Trends ในช่วงเวลาเดียวกันกับอัตราการว่างงานในแต่ละเดือน สามารถแสดงรายละเอียดของแบบจำลองได้ตามภาพที่ 4.11 ดังนี้

ภาพที่ 4.11

ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable SVR ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากภาพที่ 4.11 จะเห็นได้ว่าค่าพยากรณ์ต่ำที่สุดของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) แตกต่างกับเส้นต่ำสุดของฟังก์ชันวัตถุประสงค์ค่อนข้างมาก

4.1.3.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

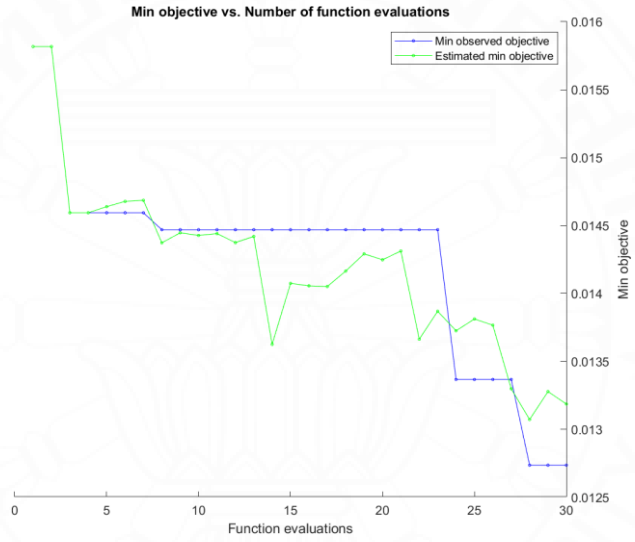
จากการนำกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) มาผ่านกระบวนการ Support Vector Regression ทำให้ได้แบบจำลองของข้อมูลชุดเรียนรู้ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยโปรแกรม MATLAB เพื่อให้ได้แบบจำลองที่

Ref. code: 25646304010116LCN

เหมาะสม และใช้ Kernel Function โดยการเพิ่ม Hyperparameter Optimization ซึ่งในการศึกษานี้จะใช้ Machine Learning ในการกำหนด ทั้งนี้ การพยากรณ์อัตราการว่างงานจากแบบจำลอง Support Vector Regression จะใช้ข้อมูลจากชุดข้อมูล Google Trends ในช่วงเวลาเดียวกันกับอัตราการว่างงานในแต่ละเดือน สามารถแสดงรายละเอียดของแบบจำลองได้ตามภาพที่ 4.12 ดังนี้

ภาพที่ 4.12

ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable SVR ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากภาพที่ 4.12 จะเห็นได้ว่าค่าพยากรณ์ต่ำที่สุดของอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) แตกต่างกับเส้นต่ำสุดของฟังก์ชันวัตถุประสงค์ค่อนข้างมาก

4.1.4 แบบจำลองโครงข่ายประสาทเทียม (Neural Network Model: ANN)

เมื่อนำชุดข้อมูล Training Data ไปสร้างแบบจำลองและใช้ชุดข้อมูล Testing Data เพื่อทดสอบแบบจำลอง โดยใช้เงื่อนไขของแบบจำลองโครงขายประสาทเทียม จากนั้นนำมาคำนวณค่าถ่วงน้ำหนัก (Optimal Weights) ของแต่ละปมของโครงข่ายให้ได้แบบจำลองที่เหมาะสมกับข้อมูลที่สุด ซึ่งจะใช้โปรแกรม MATLAB ในการสร้างแบบจำลองที่ใช้ชื่อว่า Medium Neural Network Model ได้ผลลัพธ์ในแต่ละชุดข้อมูล ดังนี้

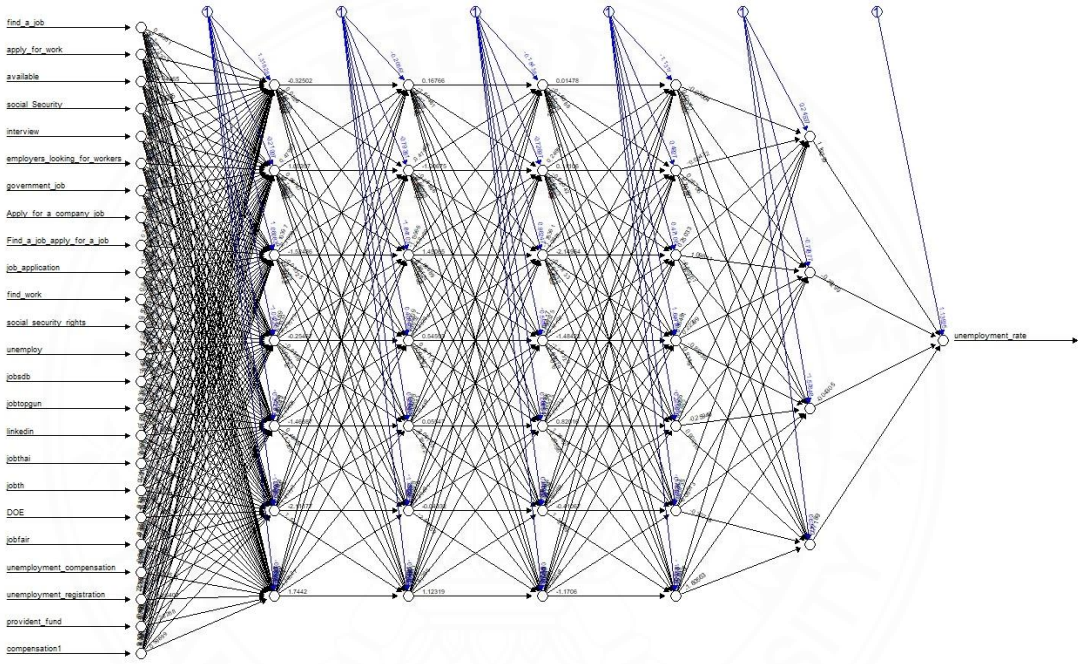
Ref. code: 25646304010116LCN

4.1.4.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)

จากการใช้อัลกอริทึมของ Machine Learning เพื่อหา Optimal Weights ให้ได้แบบจำลองที่มีความแม่นยำที่สุดในชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) สามารถแสดงรายละเอียดดังภาพที่ 4.13 ดังนี้

ภาพที่ 4.13

โครงสร้างประสาทเทียมของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา[1]

4.1.4.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

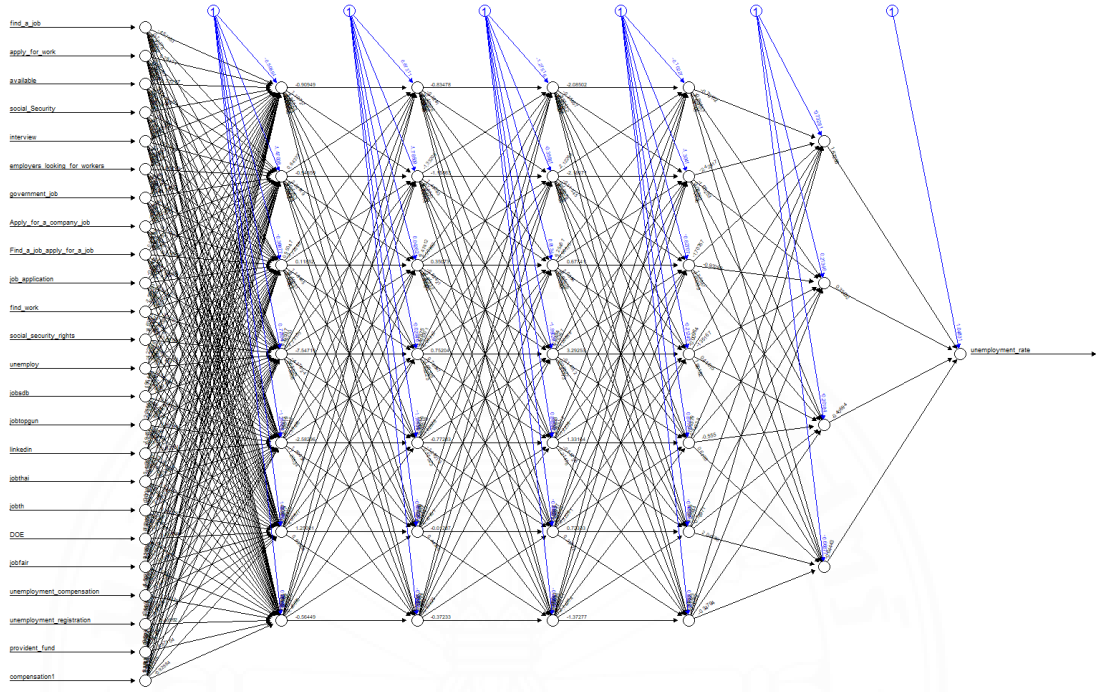
จากการใช้อัลกอริทึมของ Machine Learning เพื่อหา Optimal Weights ให้ได้แบบจำลองที่มีความแม่นยำที่สุดในชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) สามารถแสดงรายละเอียดดังภาพที่ 4.14 ดังนี้

[1] เนื่องจากโปรแกรม MATLAB ไม่สามารถสร้างผังโครงสร้างประสาทเทียมได้ ผู้ศึกษาจึงใช้โปรแกรม R Studio ในการสร้างผังโครงสร้างประสาทเทียม

Ref. code: 25646304010116LCN

ภาพที่ 4.14

โครงสร้างประสาทเทียมของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา[2]

4.1.5 แบบจำลองแบบเรียงซ้อน (Optimizable Ensemble Model: EM)

เมื่อนำชุดข้อมูล Training Data ไปสร้างแบบจำลองและใช้ชุดข้อมูล Testing Data เพื่อทดสอบแบบจำลอง โดยใช้เงื่อนไขของแบบจำลองแบบเรียงซ้อน โดยนำแบบจำลองหลาย ๆ แบบจำลองมารวมกันเพื่อที่จะทำให้ผลลัพธ์ออกมาดีที่สุด ซึ่งแบบจำลองที่นำมาใช้ จะใช้อัลกอริทึมของ Machine Learning หลายอย่าง เช่น Linear Regression, Logistic Regression, Decision Tree ขึ้นอยู่กับวิธีการใดจะสามารถสร้างแบบจำลองได้ออกมาดีที่สุด และใน Ensemble Model จะมีการนำวิธีการของ Bagging และ Boosting มาหาแบบจำลองเพื่อให้ได้ประสิทธิภาพสูงที่สุด ซึ่งจะใช้โปรแกรม MATLAB ในการสร้างแบบจำลองที่ใช้ชื่อว่า Optimizable Ensemble Model ได้ผลลัพธ์ในแต่ละชุดข้อมูล ดังนี้

[2] เนื่องจากโปรแกรม MATLAB ไม่สามารถสร้างผังโครงสร้างประสาทเทียมได้ ผู้ศึกษาจึงใช้โปรแกรม R Studio ในการสร้างผังโครงสร้างประสาทเทียม

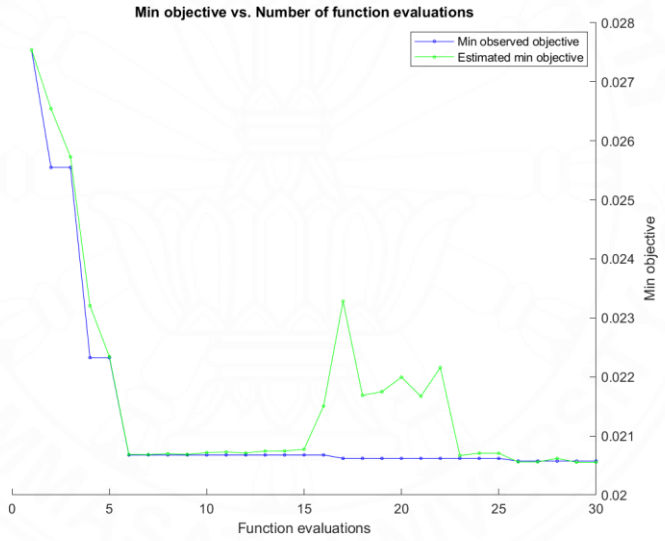
Ref. code: 25646304010116LCN

4.1.5.1 ข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)

จากการใช้อัลกอริทึมของ Machine Learning เพื่อหาแบบจำลองที่มีความแม่นยำที่สุดในชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) โดยในชุดข้อมูล Training Data ของตัวแปร Google Trends จะใช้วิธีการ Bagging ที่สามารถทำให้ผลลัพธ์ของแบบจำลองมีประสิทธิภาพที่สุด ในขณะที่ชุดข้อมูลการพยากรณ์อัตราการว่างงาน จะใช้วิธีการ Boosting ที่สามารถทำให้ผลลัพธ์ของแบบจำลองมีประสิทธิภาพที่สุด สามารถแสดงรายละเอียดได้ดังนี้

ภาพที่ 4.15

ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Ensemble ของข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากภาพที่ 4.15 จะเห็นได้ว่าค่าพยากรณ์ต่ำที่สุดของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ค่อนข้างแนบสนิทเส้นต่ำสุดของฟังก์ชันวัตถุประสงค์ แต่จะมีเพียงบางจุดที่แตกต่างกันค่อนข้างมาก

4.1.5.2 ข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)

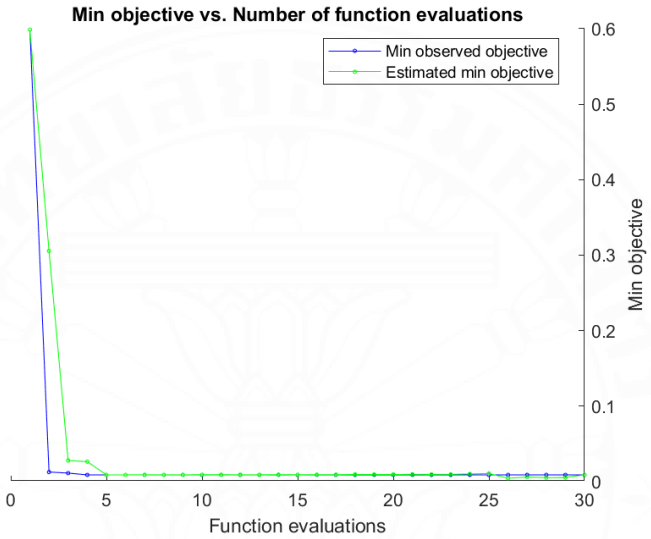
จากการใช้อัลกอริทึมของ Machine Learning เพื่อหาแบบจำลองที่มีความแม่นยำที่สุดในชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) โดยพบว่าในชุดข้อมูล Training Data ของตัวแปร Google Trends และ ชุดข้อมูลการพยากรณ์อัตราการว่างงาน

Ref. code: 25646304010116LCN

จะใช้วิธีการ Boosting ที่สามารถทำให้ผลลัพธ์ของแบบจำลองมีประสิทธิภาพที่สุด สามารถแสดงรายละเอียดได้ ดังนี้

ภาพที่ 4.16

ค่าต่ำสุดของฟังก์ชันวัตถุประสงค์ และจำนวนของการประเมินค่าฟังก์ชันจากวิธี Optimizable Ensemble ของข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO)



หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากภาพที่ 4.16 จะเห็นได้ว่าค่าพยากรณ์ต่ำที่สุดของอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ค่อนข้างแนบสนิทเส้นต่ำสุดของฟังก์ชันวัตถุประสงค์

4.2 ผลการทดสอบแบบจำลองกับชุดข้อมูลทดสอบ (Testing Data)

เพื่อเปรียบเทียบความแม่นยำของแบบจำลองในแต่ละวิธีการศึกษา ซึ่งจากการศึกษาพบว่าผลลัพท์ที่ได้ในแต่ละแบบจำลองทั้ง 5 แบบจำลอง นำมาเปรียบเทียบกับค่าอัตราการว่างงานจริง ดังนี้

Ref. code: 25646304010116LCN

4.2.1 รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE)

ในการศึกษานี้ ผู้ศึกษาจะใช้วิธีการตรวจสอบผลความแม่นยำในการพยากรณ์ของแบบจำลองด้วยการคำนวณรากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ซึ่งจะพิจารณาทั้งจากข้อมูลที่ใช้ในการสร้างแบบจำลอง (Training Data) และข้อมูลที่ใช้ทดสอบความถูกต้องของแบบจำลอง (Testing Data) เพื่อเป็นการเปรียบเทียบค่าพยากรณ์ที่คำนวณได้จากวิธีที่เลือกมากับค่าข้อมูลจริง ณ ช่วงเวลานั้น ๆ ว่าแตกต่างกันมากน้อยเพียงใด โดยหาก RMSE มีค่าน้อยจะแสดงถึงความผิดพลาดในการพยากรณ์ที่น้อยลง และสะท้อนประสิทธิภาพความแม่นยำของแบบจำลอง ทั้งนี้ สามารถแสดงค่า RMSE ได้ดังนี้

ตารางที่ 4.6

ค่าความคาดเคลื่อนรากที่สองกำลังสองเฉลี่ย (RMSE) ในแต่ละแบบจำลอง

RMSE	NSO		SSO	
	Training Data	Testing Data	Training Data	Testing Data
Linear Regression	0.163810	0.668240	0.138620	1.212900
Optimizable Tree	0.163480	0.453570	0.099682	1.275500
Optimizable SVR	0.140610	0.512000	0.094596	1.349600
Medium Neural Network	0.215470	1.473000	0.222540	1.831300
Optimizable Ensemble	0.141870	0.435090	0.092270	1.246500

หมายเหตุ. รวบรวมโดยผู้ศึกษา

จากผลการศึกษาการพยากรณ์อัตราการว่างงาน พบว่า แบบจำลองของชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) สำหรับชุดข้อมูล Training Data จะมีค่าคาดเคลื่อนของวิธีการ Optimizable SVR เท่ากับ 0.14061 ซึ่งมีค่าน้อยที่สุดเมื่อเทียบกับวิธีการอื่น ขณะที่ชุดข้อมูล Testing Data จะมีค่าคาดเคลื่อนของวิธีการ Optimizable Ensemble เท่ากับ 0.43509 ซึ่งมีค่าน้อยที่สุดเมื่อเทียบกับวิธีการอื่น แต่ในทางกลับกันแบบจำลองของชุดข้อมูลอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) สำหรับชุดข้อมูล Training Data จะมีค่าคาดเคลื่อนของวิธีการ Optimizable Ensemble เท่ากับ 0.09227 ซึ่งมีค่าน้อยที่สุดเมื่อเทียบกับวิธีการอื่น

Ref. code: 25646304010116LCN

ขณะที่ชุดข้อมูล Testing Data จะมีค่าคาดเคลื่อนของวิธีการ Linear Regression เท่ากับ 1.2129 ซึ่งมีค่าน้อยที่สุด

ทั้งนี้ จะเห็นได้ว่าชุดข้อมูล Training Data จะมีค่าคาดเคลื่อนของวิธีการ Optimizable SVR และ Optimizable Ensemble น้อยที่สุดและใกล้เคียงกันทั้ง 2 ชุดข้อมูล เนื่องจาก SVR เหมาะกับการพยากรณ์ที่มีข้อมูลขนาดเล็ก แต่มีตัวแปรค่อนข้างมาก และการสร้างแบบจำลองด้วยวิธีการ Ensemble เป็นวิธีการสร้างแบบจำลองหลาย ๆ แบบจำลองที่ไม่เหมือนกัน แล้วนำผลลัพธ์ในแต่ละแบบจำลองมารวมกัน ประกอบกับการเรียนรู้ของชุดข้อมูลที่แบบจำลองเคยเห็นข้อมูลนั้นมาก่อนในการสร้างแบบจำลองด้วยวิธีการดังกล่าวมีประสิทธิภาพที่สุด นอกจากนี้ แบบจำลองทั้ง 2 ชุดข้อมูล ได้ผลลัพธ์ของค่าคาดเคลื่อนที่แตกต่างกันในแต่ละวิธีการสร้างแบบจำลอง เป็นผลมาจากชุดข้อมูล Testing Data เป็นการเปรียบเทียบกับชุดข้อมูลที่แบบจำลองไม่เคยเรียนรู้มาก่อน ดังนั้น จึงควรพิจารณาประสิทธิภาพของแบบจำลองจากค่าคาดเคลื่อนที่ได้จากชุดข้อมูล Testing Data ซึ่งสามารถสะท้อนให้เห็นถึงประสิทธิภาพในการพยากรณ์อัตราการว่างงานที่มีความแม่นยำมากกว่าวิธีอื่น หรือมีค่าความผิดพลาด หรือค่าความคลาดเคลื่อนต่ำกว่าวิธีอื่น อย่างไรก็ตาม ค่าคาดเคลื่อนของวิธีการ Neural Network ของทั้งแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) และอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) มีค่าคลาดเคลื่อนที่สูงที่สุดเมื่อเทียบกับวิธีการอื่น

4.2.2 Scatter plot

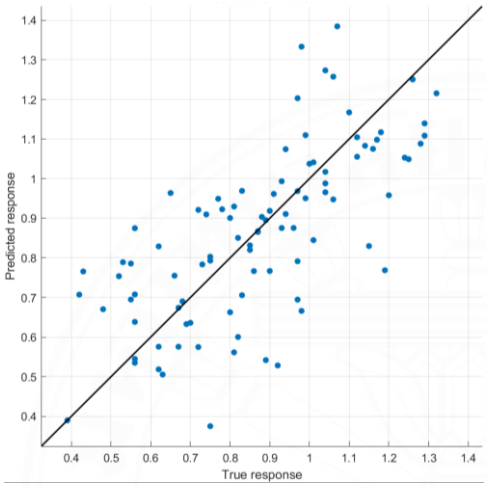
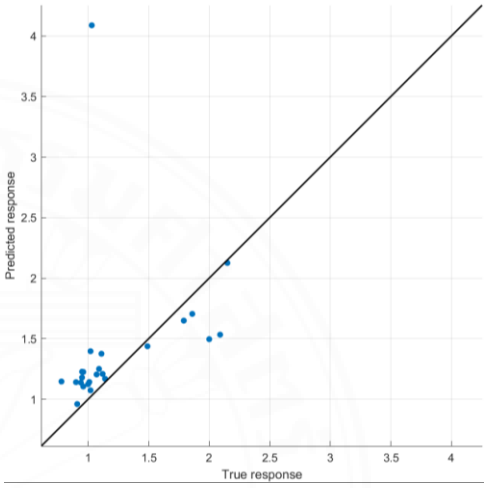
แสดงถึงตัวแปรตามหรืออัตราการว่างงานจากข้อมูลของสำนักงานสถิติแห่งชาติ เปรียบเทียบกับค่าจริงจากกลุ่มข้อมูลที่ถูกนำไปเรียนรู้เพื่อสร้างแบบจำลอง (Training Data) และเปรียบเทียบกับค่าจริงจากกลุ่มข้อมูลที่ใช้ทดสอบความถูกต้องของแบบจำลอง (Testing Data) ซึ่งหากค่าที่ได้จากแบบจำลองมีความใกล้เคียงกับค่าที่เกิดขึ้นจริงตำแหน่งบนจุดของภาพจะต้องอยู่ใกล้เคียงหรือทับเส้นกับเส้นเฉียงที่ทำมุม 45 องศา หรือเส้น Perfect Prediction

Ref. code: 25646304010116LCN

4.2.2.1 Linear Regression Model

ตารางที่ 4.7

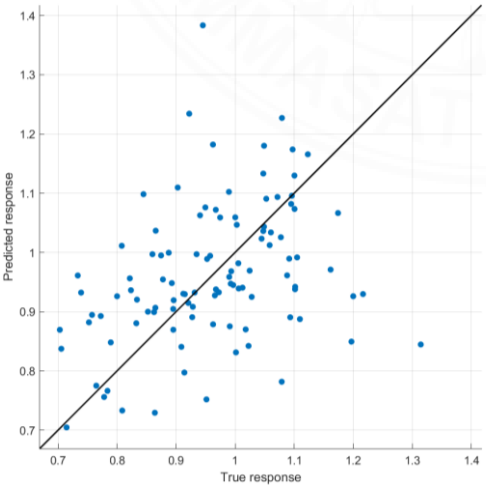
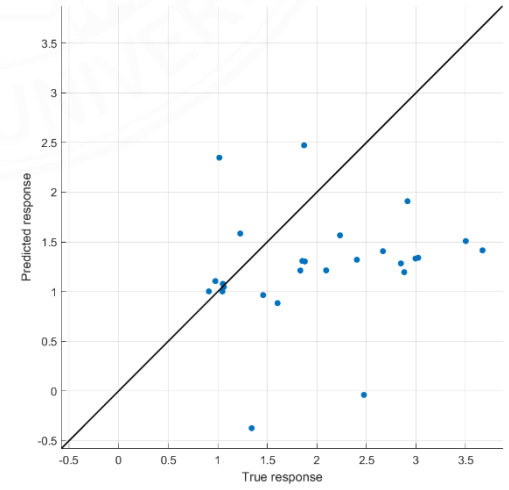
ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Linear Regression

Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ตารางที่ 4.8

ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Linear Regression

Training Data	Testing Data
	

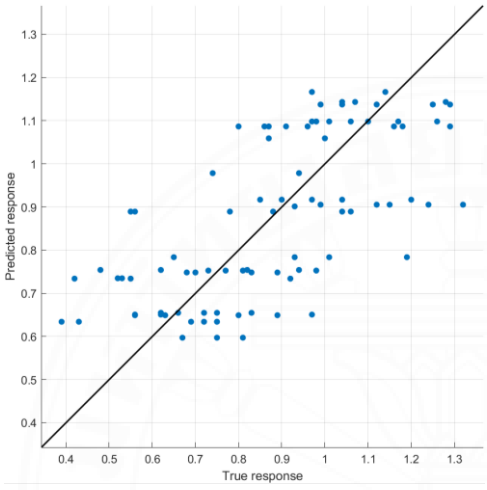
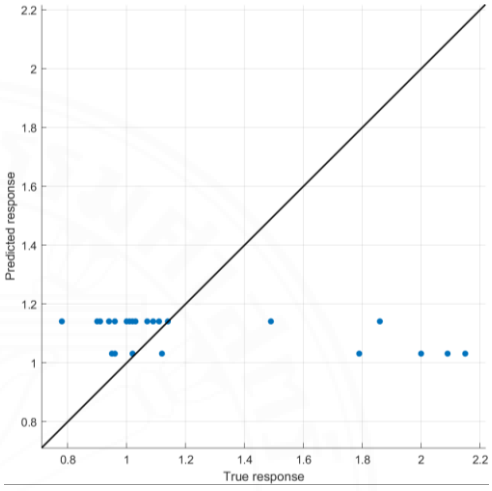
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.2.2 Optimizable Decision Tree Regression Model

ตารางที่ 4.9

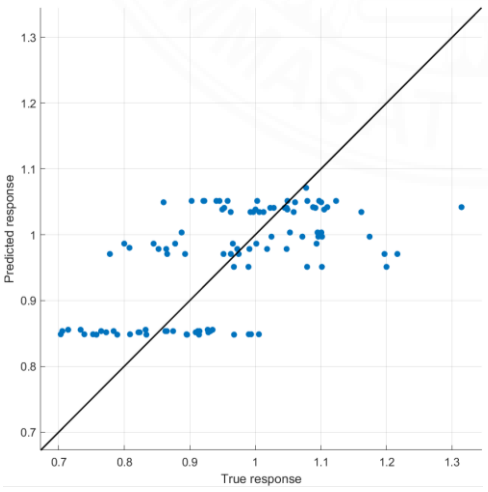
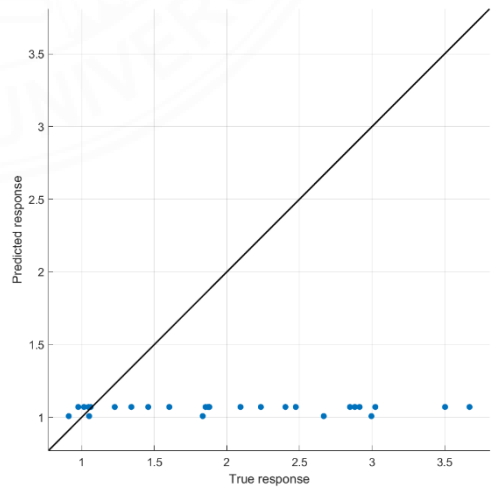
ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Optimizable Tree

Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ตารางที่ 4.10

ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Optimizable Tree

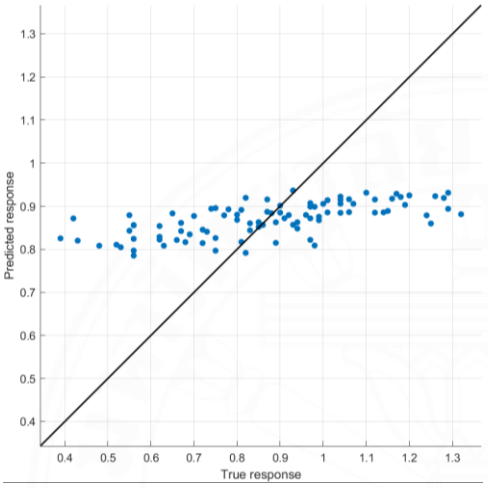
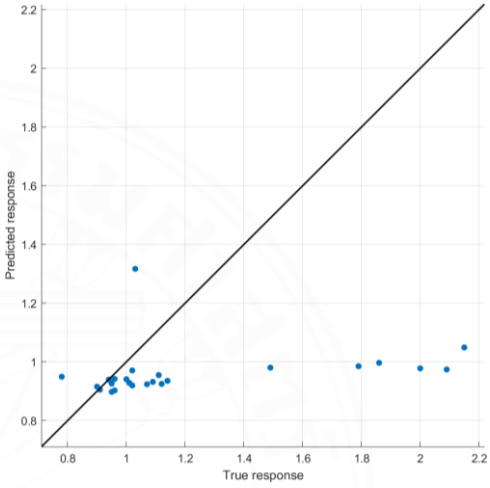
Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

4.2.2.3 Optimizable Support Vector Regression Model

ตารางที่ 4.11

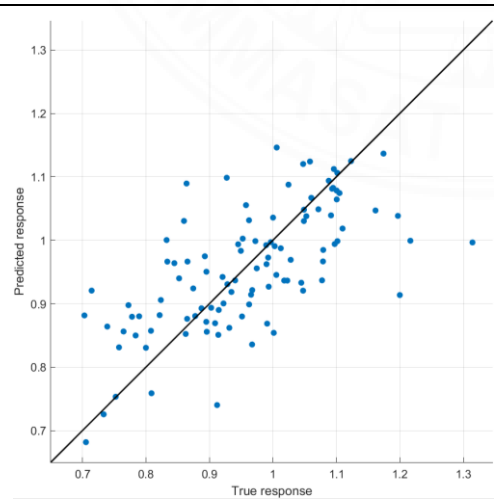
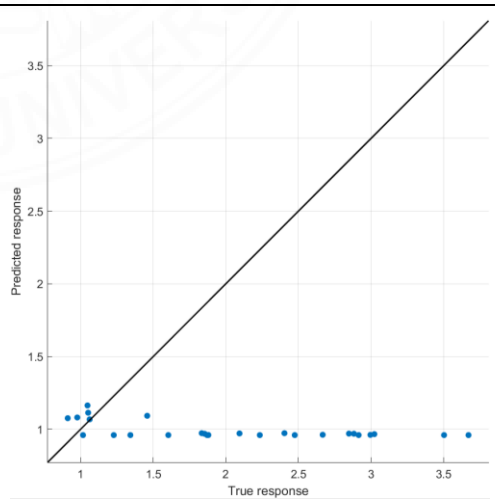
ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Optimizable SVR

Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ตารางที่ 4.12

ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Optimizable SVR

Training Data	Testing Data
	

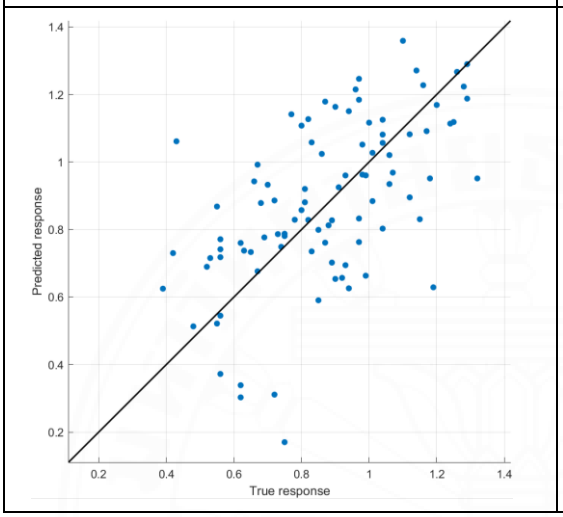
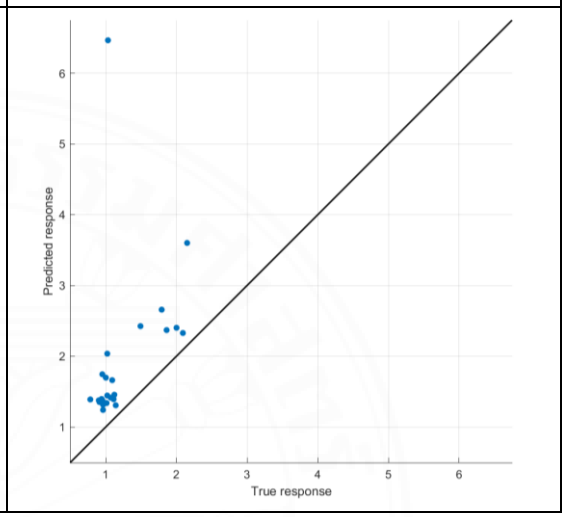
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.2.4 Neural Network Model

ตารางที่ 4.13

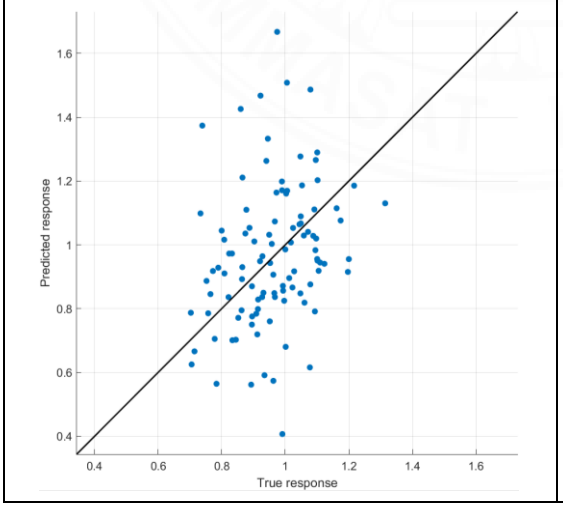
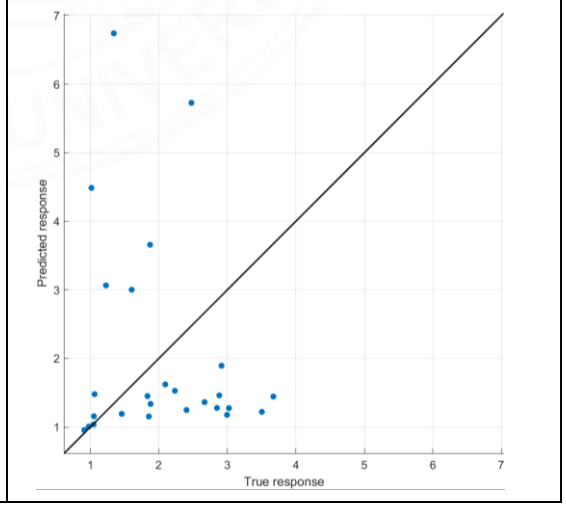
ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Neural Network

Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ตารางที่ 4.14

แสดงค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Neural Network

Training Data	Testing Data
	

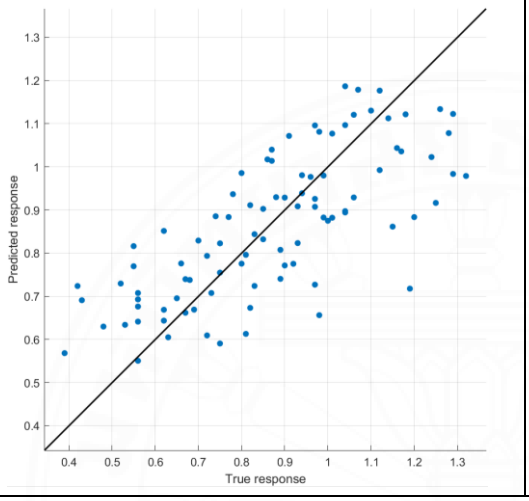
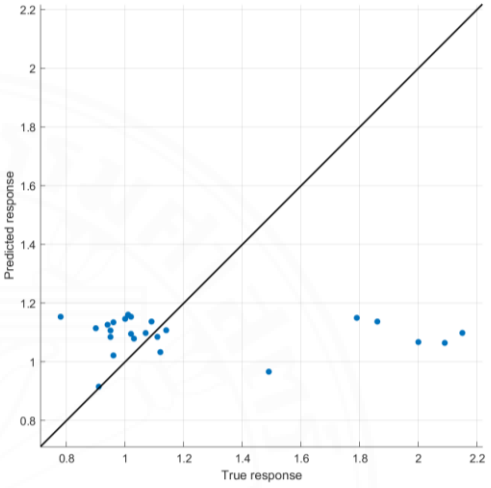
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.2.5 Optimizable Ensemble Model

ตารางที่ 4.15

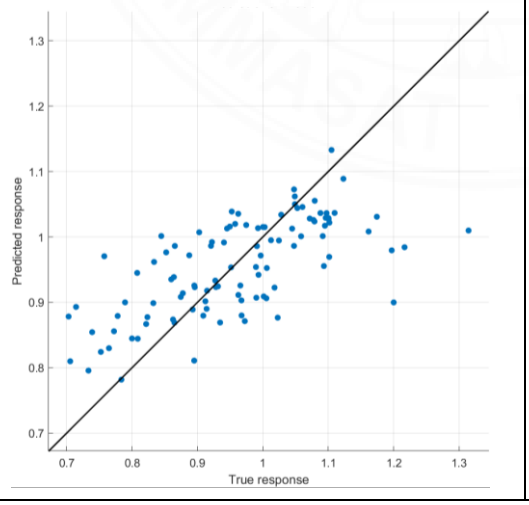
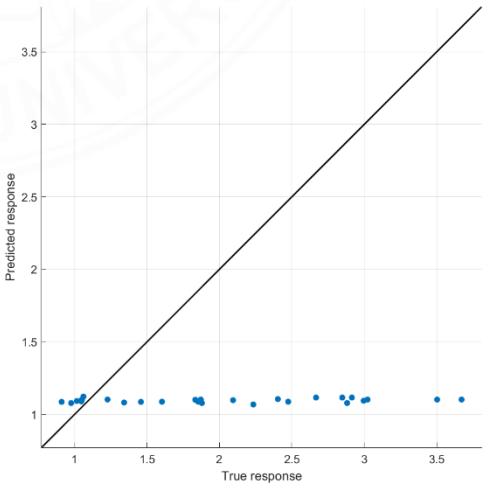
ค่า Scatter plot ของชุดข้อมูล NSO ด้วยวิธีการ Optimizable Ensemble

Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

ตารางที่ 4.16

ค่า Scatter plot ของชุดข้อมูล SSO ด้วยวิธีการ Optimizable Ensemble

Training Data	Testing Data
	

หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

จากภาพที่แสดงค่า Scatter plot ในแต่ละแบบจำลอง พบว่า ในภาพของ Training Data ในแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) วิธีการ Optimizable SVR จะมีตำแหน่งของจุดกระจุกตัวกันเป็นเส้นตรงและใกล้เคียงกับเส้นทแยงมุม 45 องศา มากกว่าวิธีอื่น ขณะที่แบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) วิธีการ Optimizable Ensemble จะมีตำแหน่งของจุดอยู่ใกล้เคียงกับเส้นทแยงมุม 45 องศา มากกว่าวิธีอื่น ซึ่งสะท้อนว่าค่าของตัวแปรที่ได้จากการการพยากรณ์ของแบบจำลอง กับค่าที่เป็นค่าจริงของข้อมูลตัวอย่างมีความใกล้เคียงกันในระดับสูงกว่าวิธีการพยากรณ์ในแบบจำลองอื่น ๆ ทั้งนี้สอดคล้องกับการวิเคราะห์ค่าความคลาดเคลื่อนรากที่สองกำลังสอง อย่างไรก็ตาม จะเห็นได้ว่าแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Neural Network จะมีตำแหน่งของจุดกระจายตัวและมีค่าผิดพลาดในการพยากรณ์มากกว่าวิธีการอื่น เนื่องจากอาจเป็นข้อจำกัดของวิธีการวิเคราะห์ข้อมูลตัวอย่างในการจัดการคำนวณกับชุดของข้อมูลที่มีลักษณะเป็นตัวแปรหุ่นจำนวนมาก ทำให้การประมวลผลอาจจะผิดพลาดมากขึ้น

ในส่วนของภาพของ Testing Data พบว่า ในแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) วิธีการ Optimizable Ensemble จะมีตำแหน่งจุดที่กระจุกตัวกับเส้นทแยงมุม 45 องศา มากกว่าวิธีอื่น ถึงแม้ว่าจะมีจุดที่คลาดเคลื่อนออกจากเส้นทแยงมุมค่อนข้างมาก ในขณะที่แบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ส่วนมากจะมีตำแหน่งของจุดกระจายตัวที่ค่อนข้างเป็นเส้นตรงขนานกับแนวนอน เป็นผลจากการประมวลผลของแบบจำลองที่ผิดพลาด และข้อมูลที่ใช้ใน Testing Data เป็นช่วงเวลาที่เกิดผลกระทบจากโควิด-19 ซึ่งมีลักษณะเหมือน Shock ต่อระบบเศรษฐกิจ

4.2.3 ผลการเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์

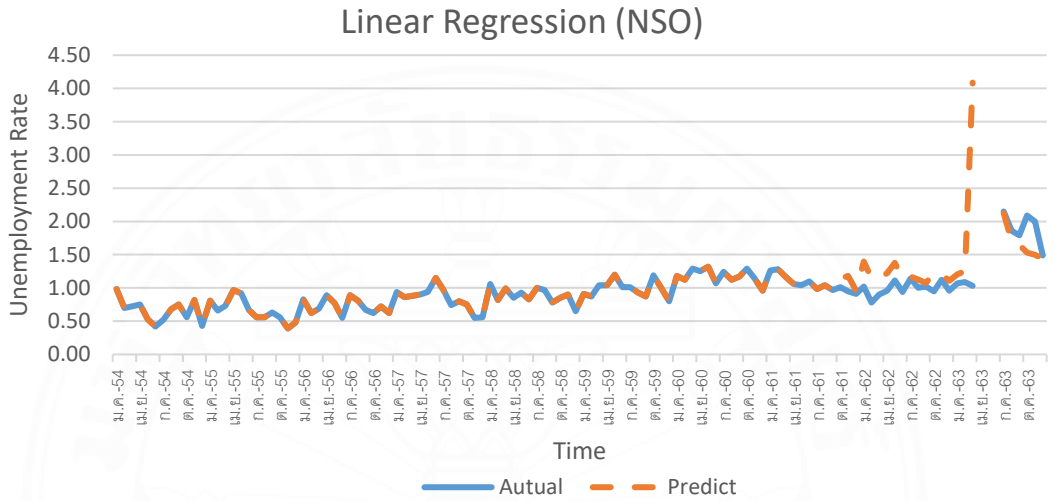
เมื่อนำค่าพยากรณ์ที่ได้จากชุดข้อมูล Training Data มาเปรียบเทียบกับชุดข้อมูล Testing Data ซึ่งเป็นค่าจริงที่เก็บรวบรวมได้ สามารถแสดงได้ ดังนี้

Ref. code: 25646304010116LCN

4.2.3.1 Linear Regression Model

ภาพที่ 4.17

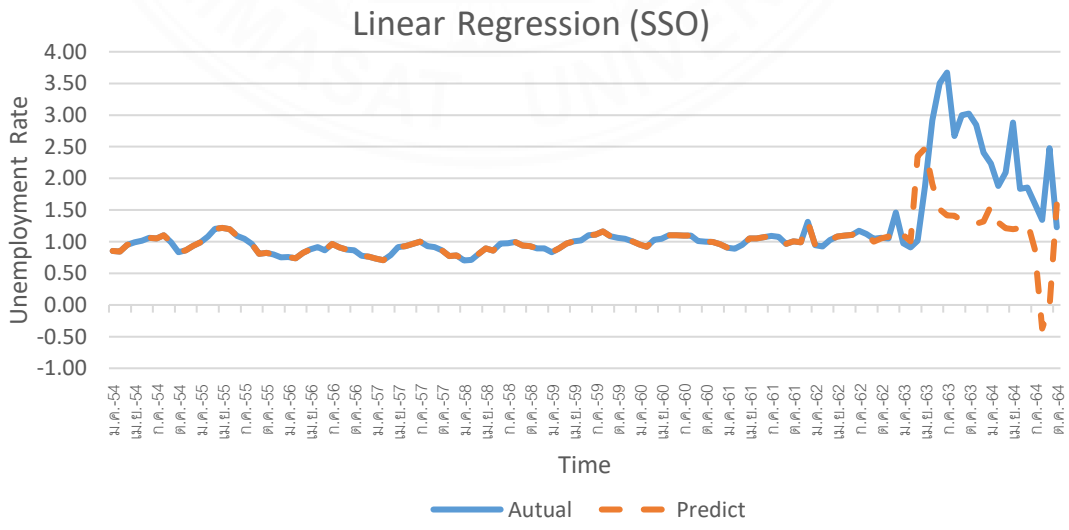
การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Linear Regression



หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.18

การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Linear Regression



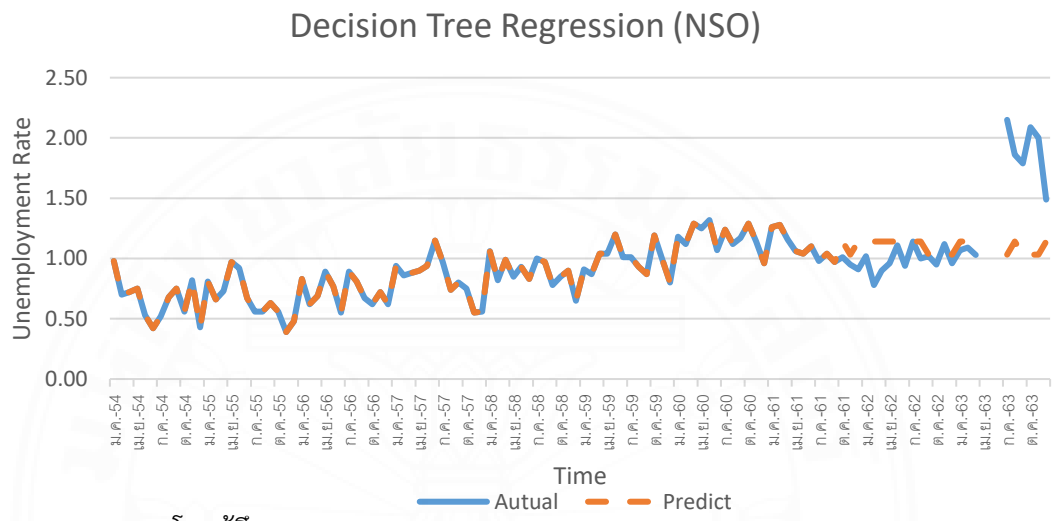
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.3.2 Optimizable Decision Tree Regression Model

ภาพที่ 4.19

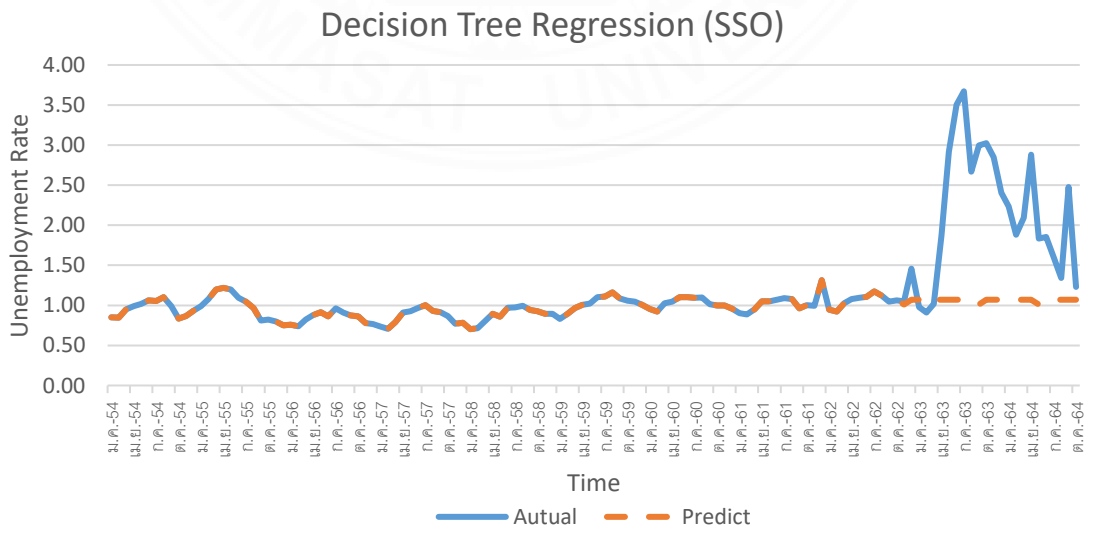
การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Decision Tree Regression



หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.20

การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Decision Tree Regression



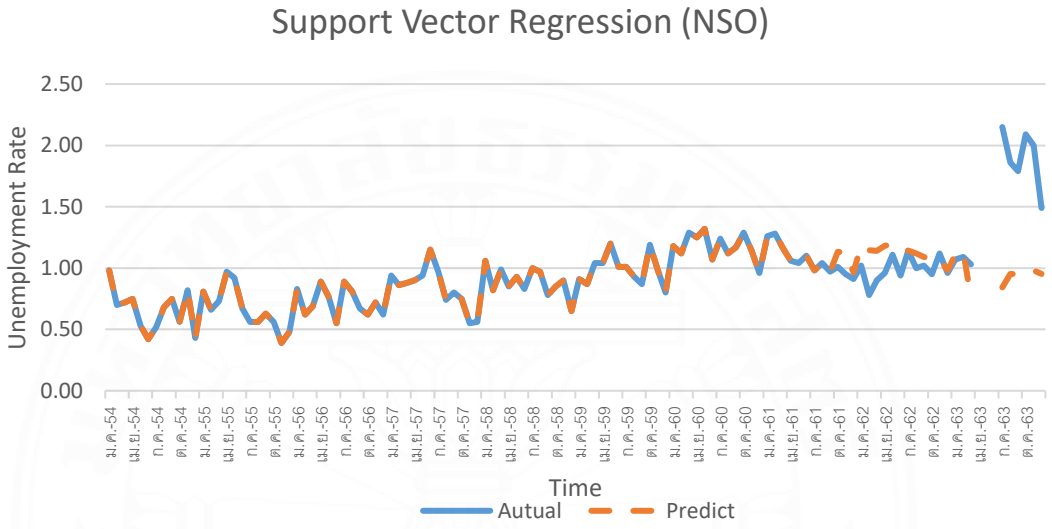
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.3.3 Optimizable Support Vector Regression Model

ภาพที่ 4.21

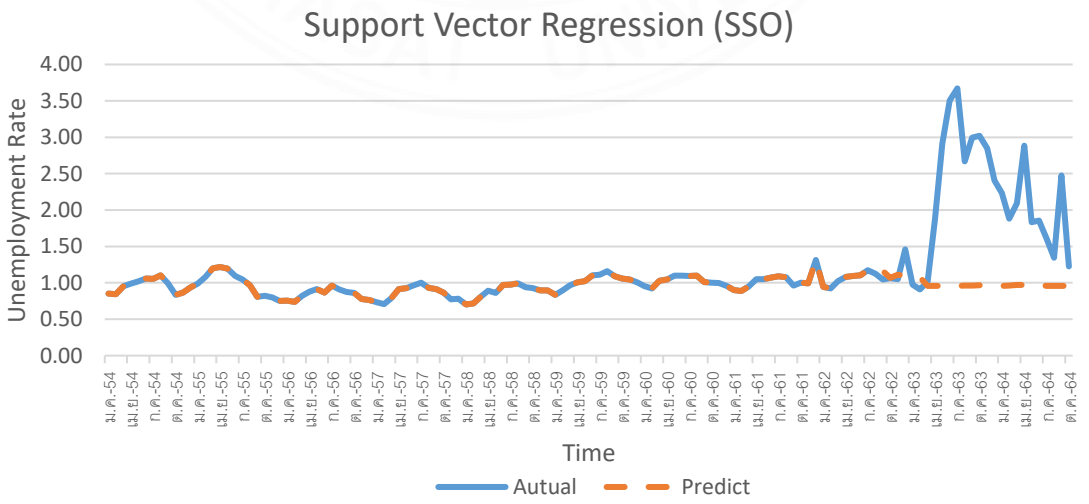
การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Support Vector Regression



หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.22

การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Support Vector Regression



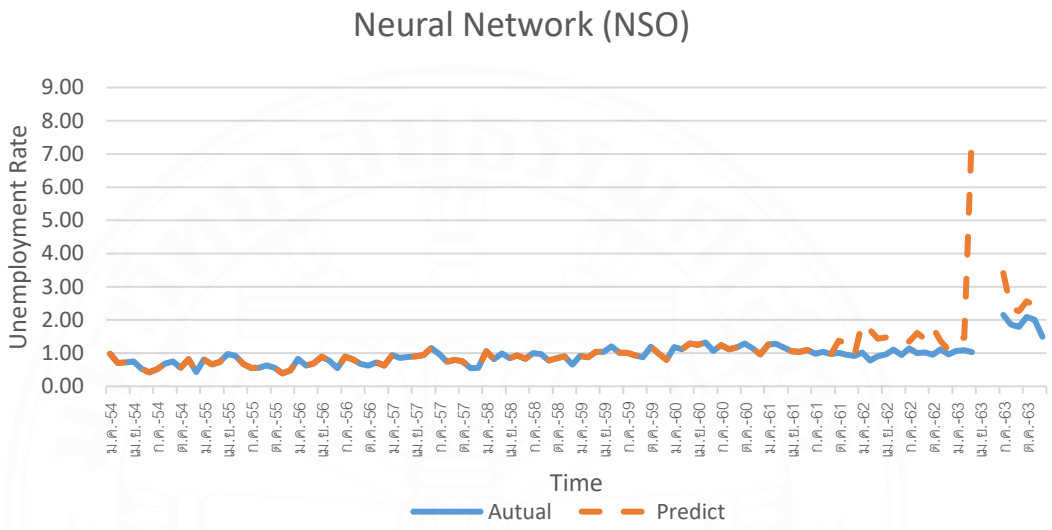
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.3.4 Neural Network Model

ภาพที่ 4.23

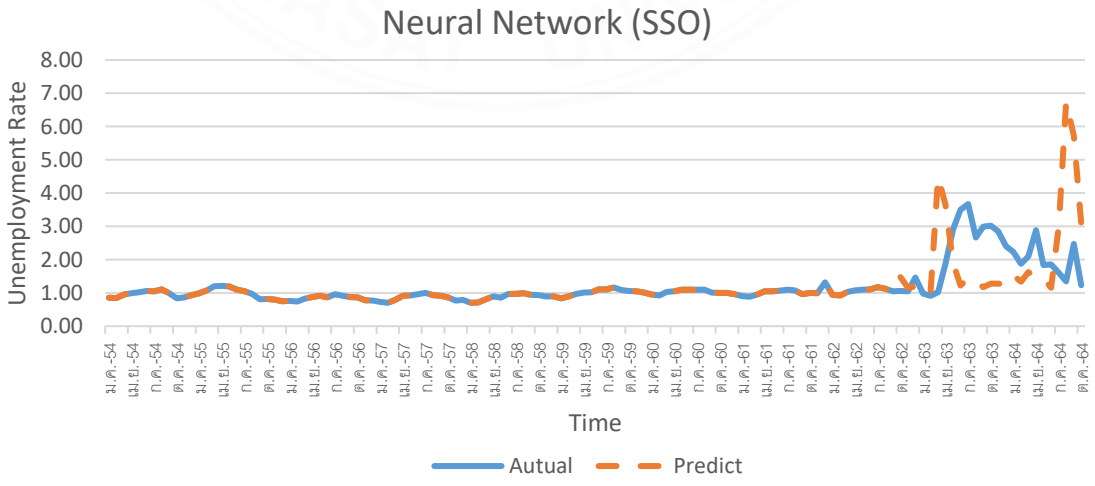
การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Neural Network



หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.24

การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Neural Network



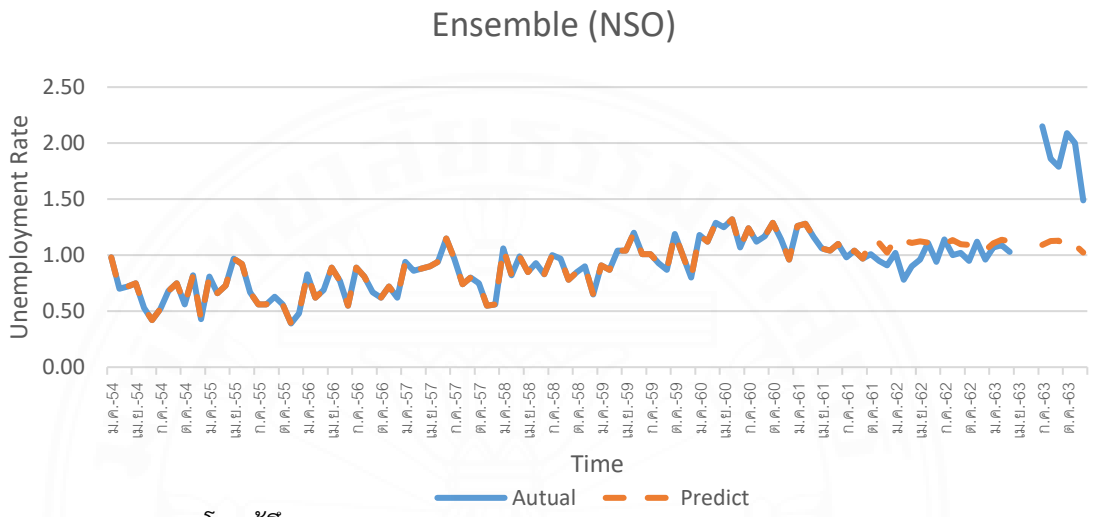
หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

4.2.3.5 Optimizable Ensemble Model

ภาพที่ 4.25

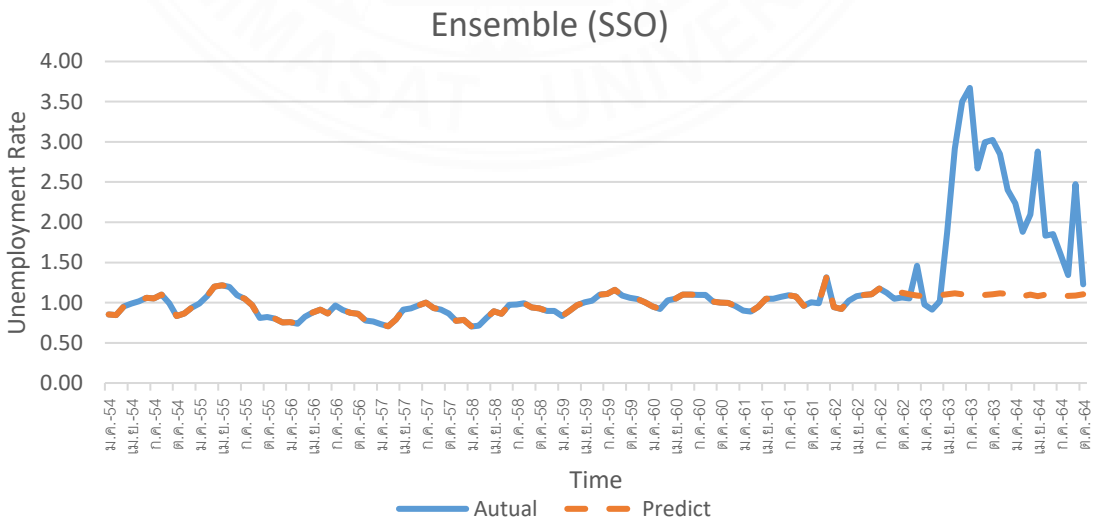
การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Ensemble



หมายเหตุ. รวบรวมโดยผู้ศึกษา

ภาพที่ 4.26

การเปรียบเทียบอัตราการว่างงานระหว่างค่าจริงและค่าพยากรณ์ของแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Ensemble



หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

จากภาพที่ 4.17 – 4.26 พบว่า ในทุก ๆ แบบจำลองจะมีการพยากรณ์ที่คลาดเคลื่อนกว่าค่าจริงค่อนข้างมากในช่วงตั้งแต่ปีพ.ศ. 2563 เนื่องจากเกิดการแพร่ระบาดของเชื้อโควิด-19 ซึ่งเป็นปัจจัยภายนอกที่เข้ามากระทบข้อมูล (Shock) ส่งผลให้เกิดความคลาดเคลื่อนระหว่างค่าอัตราการว่างงานจริงกับค่าพยากรณ์สูงในช่วงเวลาดังกล่าว แต่อย่างไรก็ตาม เมื่อพิจารณาค่าอัตราการว่างงานจริงเปรียบเทียบกับค่าพยากรณ์ในแต่ละวิธีการแล้ว จะเห็นได้ว่า แบบจำลองอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (NSO) ด้วยวิธีการ Ensemble จะมีเส้นค่าอัตราการว่างงานจริงกับเส้นค่าพยากรณ์ใกล้เคียงกันมากกว่าวิธีการพยากรณ์อื่น ถึงแม้ว่าจะมีค่าพยากรณ์ที่ต่างจากค่าจริงจำนวนหนึ่ง และแบบจำลองอัตราการว่างงานของผู้ประกันตนในประกันสังคม (SSO) ด้วยวิธีการ Linear Regression จะมีเส้นค่าอัตราการว่างงานจริงกับเส้นค่าพยากรณ์ใกล้เคียงกันมากกว่าวิธีการพยากรณ์อื่นเช่นกัน ถึงแม้ว่าจะมีค่าพยากรณ์บางค่าที่ติดลบ ซึ่งสอดคล้องกับค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE)

Ref. code: 25646304010116LCN

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การศึกษาเรื่อง “การประยุกต์ใช้ Google Trends ในการประมาณการอัตราการว่างงานในประเทศไทย” เพื่อศึกษาการใช้ข้อมูลคำค้นหาที่เกี่ยวข้องกับการหางานใน Google Trends ในการประมาณการอัตราการว่างงานในอนาคต โดยการสร้างแบบจำลองด้วยวิธีการของ Machine Learning ซึ่งเป็นวิธีทางเลือกสมัยใหม่ ได้ข้อมูลแบบทันการณ์ และเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองในแต่ละวิธีการของ Machine Learning ตลอดจนสามารถใช้ในการพยากรณ์ตัวเลขทางเศรษฐกิจในกรณีที่ประเทศเกิดเหตุการณ์ฉุกเฉินทำให้ไม่สามารถเก็บข้อมูลทางสถิติได้ครบถ้วน ทั้งนี้ ผู้ศึกษาได้ทำการสรุปผลการศึกษาและข้อเสนอแนะ ดังนี้

5.1 การสรุปและอภิปรายผลการศึกษา

จากผลการศึกษาการสร้างแบบจำลองการพยากรณ์อัตราการว่างงานด้วยวิธีการ Machine Learning ทั้ง 5 วิธีการ เมื่อพิจารณาจากการนำข้อมูล Testing Data มาทดสอบกับแบบจำลองที่สร้างจาก Training Data เนื่องจากเป็นการทดสอบการเรียนรู้ของแบบจำลองกับข้อมูลที่แบบจำลองในแต่ละวิธีการไม่เคยเห็นมาก่อน พบว่า แบบจำลองชุดข้อมูลอัตราการว่างงานของสำนักงานสถิติแห่งชาติ (National Statistical Office: NSO) ด้วยวิธีการ Ensemble มีประสิทธิภาพสูงสุด ในขณะที่แบบจำลองชุดข้อมูลอัตราการว่างงานของสำนักงานประกันสังคม (Social Security Office: SSO) ด้วยวิธีการ Linear Regression มีประสิทธิภาพสูงสุด ซึ่งจะพิจารณาจากค่าความคลาดเคลื่อนรากที่สองกำลังสองเฉลี่ย (RMSE) ที่ต่ำที่สุด และผลจากภาพ Scatter plot ที่กระจุกตัวใกล้กับเส้นทแยงมุม 45 องศา อย่างไรก็ตาม Ensemble ไม่สามารถอธิบายถึงปัจจัยที่มีผลต่ออัตราการว่างงานได้ แต่สามารถทำได้เพียงพยากรณ์ผลลัพธ์ที่มีความแม่นยำ แต่ Linear Regression สามารถระบุถึงตัวแปรที่มีอิทธิพลกับอัตราการว่างงาน ซึ่งประกอบด้วยคำว่า “หางานสมัครงาน”, “งานราชการ”, “JobDB”, “JobTopGun” และ “ลงทะเบียนว่างงาน” ทั้งนี้เมื่อใช้ข้อมูลทั้ง 24 ตัวแปร มาใช้ร่วมกันในแบบจำลองเพื่อการพยากรณ์จะทำให้แบบจำลองอัตราการว่างงานของสำนักงานประกันสังคม (SSO) ทำให้สามารถพยากรณ์ได้มีประสิทธิภาพมากขึ้น สามารถสรุปผลประสิทธิภาพของแต่ละแบบจำลองได้ ดังนี้

Ref. code: 25646304010116LCN

ตารางที่ 5.1

สรุปผลประสิทธิภาพของแบบจำลอง

	RMSE				Scatter plot			
	NSO		SSO		NSO		SSO	
	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data	Training Data	Testing Data
Linear Regression	มาก	มาก	มาก	น้อยที่สุด	กระจายตัวตามเส้น 45 องศา	กระจุกตัว, มีค่าผิดปกติ	กระจายตัวตามเส้น 45 องศา,มีค่าผิดปกติ	กระจายตัวตามเส้น 45 องศา,มีค่าผิดปกติ
Decision Tree	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	กระจายตัวตามเส้น 45 องศา	กระจายตัวขนานกับแกนนอน	กระจายตัวขนานกับแกนนอน	กระจายตัวขนานกับแกนนอน
SVR	น้อยที่สุด	น้อย	น้อย	มาก	กระจายตัวตามเส้น 45 องศา	กระจุกตัว, มีค่าผิดปกติ	กระจายตัวตามเส้น 45 องศา	กระจายตัวขนานกับแกนนอน
Neural Network	มากที่สุด	มากที่สุด	มากที่สุด	มากที่สุด	กระจายตัวตามเส้น 45 องศา, มีค่าผิดปกติ	กระจุกตัว, มีค่าผิดปกติ	กระจุกตัวเป็นกลุ่ม, มีค่าผิดปกติ	กระจายตัวสูง, มีค่าผิดปกติ
Ensemble	น้อย	น้อยที่สุด	น้อยที่สุด	น้อย	กระจายตัวตามเส้น 45 องศา	กระจุกตัว, มีค่าผิดปกติ	ค่อนข้างกระจุกตัวกับเส้น 45 องศา	กระจายตัวขนานกับแกนนอน

หมายเหตุ. รวบรวมโดยผู้ศึกษา

Ref. code: 25646304010116LCN

5.2 ข้อจำกัดของงานวิจัย และข้อเสนอแนะ

ในการศึกษาครั้งนี้ เป็นการศึกษาจากข้อมูลอัตราการว่างงานที่รวบรวมมาจาก 2 หน่วยงานภาครัฐของประเทศไทย ซึ่งแต่ละแหล่งที่มาของข้อมูลมีข้อจำกัดของลักษณะของข้อมูลที่แตกต่างกัน กล่าวคือ

1. ข้อมูลอัตราการว่างงานจากการรวบรวมข้อมูลการสำรวจประชากรของสำนักงานสถิติแห่งชาติ (National Statistical Office: NSO) โดยข้อมูลที่นำมาเผยแพร่ต่อสาธารณะเริ่มในปี พ.ศ. 2554 และมีความถี่เป็นแบบรายเดือน ทำให้มีข้อมูลไม่มากพอที่จะใช้ในการสร้างแบบจำลอง (Training Data) ซึ่งอาจส่งผลให้การสร้างแบบจำลองเกิดความไม่แม่นยำเท่าที่ควร อีกทั้งในช่วงเดือนเมษายน - มิถุนายน ปี พ.ศ. 2563 ไม่พบการเก็บข้อมูลอัตราการว่างงาน เนื่องจากสถานการณ์การแพร่ระบาดของเชื้อไวรัสโควิด-19 ที่ไม่เคยเกิดขึ้นมาก่อน ทำให้ไม่สามารถสำรวจการว่างงานของประชากรได้ ซึ่งส่งผลต่อการสร้างแบบจำลอง เนื่องจากข้อมูลอัตราการว่างงานมีลักษณะการเคลื่อนไหวของข้อมูลเป็นวัฏจักรรายเดือน การที่ขาดข้อมูลช่วงไตรมาส 2 ปี พ.ศ. 2563 อาจจะส่งผลต่อประสิทธิภาพในการสร้างแบบจำลอง

2. ข้อมูลอัตราการว่างงานจากการรวบรวมข้อมูลของผู้ลงทะเบียนประกันตนในประกันสังคมที่ขอรับผลประโยชน์ทดแทนในกรณีว่างงาน (Social Security Office: SSO) มีข้อจำกัดในการเก็บข้อมูลที่ใช้ในการสร้างแบบจำลอง (Training Data) มีจำนวนไม่มากพอ และเป็นเพียงจำนวนหนึ่งของประชากรไทย รวมถึงกลุ่มที่ลงทะเบียนมีการกระจุกตัว เนื่องจากมีผู้ที่ไม่ได้เข้าระบบประกันสังคมอีกเป็นจำนวนมาก จากการไม่สามารถเข้าถึงข้อมูลข่าวสาร หรือด้วยเหตุผลต่าง ๆ ข้อมูลเหล่านี้จึงอาจไม่สะท้อนการว่างงานของประชากรไทยทั้งหมด

อีกทั้ง ข้อมูลอัตราการว่างงานดังกล่าวมีข้อมูลที่ใช้ในการทดสอบแบบจำลอง (Testing Data) เป็นช่วงของข้อมูลที่เกิดความผันผวนของอัตราการว่างงานสูง เนื่องจากก่อนเกิดสถานการณ์โควิด-19 อัตราการว่างงานในประเทศไทยค่อนข้างทรงตัว เมื่อเป็นเช่นนี้ทำให้เมื่อนำข้อมูล Testing Data มาทดสอบจึงเกิดค่าความคลาดเคลื่อนที่ค่อนข้างสูง

ดังนั้น สำหรับการศึกษาในครั้งถัดไป การสร้างแบบจำลองด้วย Machine Learning ควรพิจารณาถึงลักษณะและคุณสมบัติของข้อมูล รวมถึงเลือกวิธีการสร้างแบบจำลองที่เหมาะสม ตลอดจนพยายามหลีกเลี่ยงการใช้ข้อมูลที่รวบรวมมาในช่วงสถานการณ์ที่ไม่ปกติ เพื่อให้งานศึกษาครั้งถัดไปสามารถพยากรณ์ได้อย่างมีประสิทธิภาพ สมบูรณ์และครบถ้วนเพิ่มขึ้น

Ref. code: 25646304010116LCN

5.3 การศึกษาต่อยอดในอนาคต

การสร้างแบบจำลองด้วยวิธีการสมัยใหม่อย่าง Machine Learning และการนำข้อมูลคำค้นหาที่เกี่ยวข้องกับค่าที่ต้องการพยากรณ์ใน Google Trends มาประยุกต์ใช้ในการพยากรณ์ข้อมูลทางเศรษฐกิจ เพื่อให้สามารถสร้างแบบจำลองได้อย่างมีประสิทธิภาพ และทันการณ์ สามารถนำไปปรับปรุงวิธีในการพยากรณ์ในอนาคตให้มีความหลากหลาย และทันสมัยมากขึ้น รวมถึงเปรียบเทียบประสิทธิภาพของแบบจำลองในอนาคตได้ ตลอดจนยังมีการใช้แบบจำลอง Machine Learning ประเภท Regression ซึ่งเหมาะสมกับข้อมูลที่มีลักษณะเป็นค่าแบบต่อเนื่อง (Continuous) ยังไม่ค่อยแพร่หลายในประเทศไทย จึงคาดว่าผู้ที่ต้องการนำงานศึกษานี้ไปต่อยอดในอนาคต สามารถนำวิธีการศึกษาไปใช้ในการพยากรณ์ทางเศรษฐกิจที่มีลักษณะของข้อมูลเป็นตัวเลขแบบต่อเนื่อง หรือนำไปพยากรณ์ข้อมูลอื่น ๆ ที่มีลักษณะเช่นเดียวกันได้อย่างเหมาะสม นอกจากนี้ เนื่องจากข้อมูลที่นำมาใช้ในการสร้างแบบจำลองเป็นชุดข้อมูลที่มีการเปลี่ยนแปลงมากในช่วงโควิด-19 ซึ่งมีลักษณะของปัจจัยที่ไม่สามารถคาดการณ์ได้ (Shock หรือ Structural break) เข้ามากระทบข้อมูล ดังนั้น การแบ่งชุดข้อมูล Training Data และ Testing Data โดยใช้ Testing Data ในช่วงที่ได้รับผลกระทบจากโควิด-19 จึงทำให้เกิดค่าคาดเคลื่อน (Error) ในการพยากรณ์มาก อย่างไรก็ตาม อาจจะเป็นแนวทางในการศึกษาในอนาคตโดยการนำชุดข้อมูลในช่วงเวลาที่ได้รับผลกระทบของ Shock ออกจากการสร้างแบบจำลอง อีกทั้ง ควรพิจารณาถึงความสัมพันธ์ในช่วงเวลาที่ต่างกัน (Lag) เพื่อพิจารณาความสัมพันธ์ที่เกิดขึ้น ซึ่งอาจเกิดจากค่าความล่าช้า (Lagged Value) ของตัวแปรอัตราการว่างงานที่ตอบสนองต่อ Google Trends และอาจจะช่วยทำให้การพยากรณ์มีประสิทธิภาพมากขึ้น

Ref. code: 25646304010116LCN

รายการอ้างอิง

บทความวารสาร

กชพรรณ สัสเลขนันท์. (2562). จับชีพจร E-COMMERCE ด้วยข้อมูล GOOGLE SEARCH. *Focus and Quick*, 160, 1-9.

กริช สมกันธา, วิไลพร กุลตังวัฒนา และ วรวิทย์ กุลตังวัฒนา. (2556). การพัฒนาระบบประเมินบุคลากรโดยใช้โครงข่ายประสาทเทียม.*วารสารเทคโนโลยีสารสนเทศ*, 9(1), 62-63.

เดช ธรรมศิริ และ พยุง มีสัจ. (2554). การจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจเทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม. *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, 299-302.

เทอดศักดิ์ เงินมูล, พิเชษฐ เหมยคำ, วิโรจน์ ปงลังกา และวิวัฒน์ ทิพจร. (2560). การคัดแยกความสุกสตรอเบอรี่ด้วยซัพพอร์ตเวกเตอร์แมชชีน. *Naresuan University Engineering Journal*, 12(2), 58-59.

ประภัสสร แสวงสุขสันต์. (2561). การติดตามภาวะเศรษฐกิจด้วยข้อมูลปริมาณการ Search จาก Google. *Focus and Quick*, 125, 1-8.

ปรเมษฐ์ ธันวานนท์, ชัยกร ยิ่งเสรี,วรพล พงษ์เพ็ชร และธนภัทร ฆังคะจิตร. (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *Journal of Information Science and Technology*, 7(1), 16-18.

รัชภูมิ ใจกลา,พันธุทิพย์ นนทรีย และอรรถชัย จินตะเวช. (2550). การทานายผลผลิตข้าวโดยวิธีซัพพอร์ตเวกเตอรแมชีน. *รายงานการประชุมวิชาการ ศวพก. ป 2550*, 11-14.

สำนักงานประกันสังคม. (2558). ผู้ประกันตนประกันสังคม. *คู่มือผู้ประกันตน.* 9-35.

วิทยานิพนธ์

กานต์สินี เจริญกิจวัชรชัย (2562). *ปัญญาประดิษฐ์กับการพยากรณ์ค่าจ้างแรงงานไทย.* [การค้นคว้าอิสระปริญญามหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์] คลังทรัพยากรสารสนเทศอิเล็กทรอนิกส์ของมหาวิทยาลัยธรรมศาสตร์. http://ethesisarchive.library.tu.ac.th/thesis/2018/TU_2018_6004010374_9941_10208.pdf

Ref. code: 25646304010116LCN

จิราภรณ์ เจริญยิ่ง (2563). *การพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูลโดยใช้ Rapid Miner.* [การค้นคว้าอิสระปริญญาบัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ] http://ir-ithesis.swu.ac.th/dspace/bitstream/123456789/1231/1/ gs591130025.pdf

ธนสิน ถนอมพงษ์พันธ์ (2563). การพยากรณ์ยอดขายรถยนต์ในประเทศไทยด้วยคำค้นหาใน Google. [ด้วยทุนสนับสนุนการวิจัย, มหาวิทยาลัยเกษตรศาสตร์] https://research.ku.ac.th/forest/Present.aspx?PresentID=31843

ธนัตถ์พราว ลีลารัตน์ (2562). *การเปรียบเทียบความสัมพันธ์ระหว่างอัตราการว่างงานไทยกับข้อมูลที่เกี่ยวข้องบนอินเทอร์เน็ตด้วยกูเกิลเทร์นดส์ (Google Trends) แ ล ะ ท วิ ต เตอร์(Twitter).* [การค้นคว้าอิสระปริญญามหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์] คลังทรัพยากรสารสนเทศอิเล็กทรอนิกส์ของมหาวิทยาลัยธรรมศาสตร์. https://www.researchgate.net/publication/355358243_withihimnikarphyakrnxatrakarwangnganthiy_dwykhxmulcak_Google_Trends_laea_Twitter_Prediciting_the_unemployment_rate_in_Thailand_using_data_obtained_from_Google_Trends_and_Twitter

รณชัย ชื่นธวัช,กิตติศักดิ์ เกิดประสพ และนิตยาเกิดประสพ (2560). *การพยากรณ์ความต้องการใช้งานหน่วยจำหน่ายไฟฟ้าด้วยซัพพอรต์เวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน.* [การค้นคว้าอิสระปริญญามหาบัณฑิต, มหาวิทยาลัยเทคโนโลยี*สุรนารี] https://li01.tci-thaijo.org/index.php/sci_ubu/article/view/86469*

รักถิ่น เหลาหา (2560). *การพยากรณ์ความเสี่ยงการเกิดโรคเบาหวานโดยเทคนิคการทำเหมืองข้อมูล : กรณีโรงพยาบาลมหาสารคาม.* [ด้วยทุนสนับสนุนการวิจัย, มหาวิทยาลัยราชภัฏมหาสารคาม] http://research.rmu.ac.th/rdi-mis//upload/fullreport/1615219126.pdf

วรรษมนต์ เกิดจงรักษ์ (2560). *การพยากรณ์ราคาเกลือสมุทร.* [การค้นคว้าอิสระปริญญามหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์] คลังทรัพยากรสารสนเทศอิเล็กทรอนิกส์ของมหาวิทยาลัยธรรมศาสตร์. http://ethesisarchive.library.tu.ac.th/thesis/2017/TU_2017_5904010146_8244_9261.pdf

Ref. code: 25646304010116LCN

อรพิชา สังขมณีนาคร (2561). *การใช้ Google Trends ใน การพยากรณ์ตลาดเงิน (SET50).* [การค้นคว้าอิสระปริญญามหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์] คลังทรัพยากรสารสนเทศอิเล็กทรอนิกส์ของมหาวิทยาลัยธรรมศาสตร์. http://ethesisarchive.library.tu.ac.th/thesis/2018/TU_2018_6002113055_10019_10202.pdf

อัปษรศรี ม่วงคง (2552). *ปัจจัยที่มีอิทธิพลต่อพฤติกรรมการใช้บริการ และความพึงพอใจโดยรวมของลูกค้าที่มาใช้บริการภัตตาคารลอยน้ำเรือริเวอร์ไซด์.* [ปริญญานิพนธ์ปริญญามหาบัณ, มหาวิทยาลัยศรีนครินทรวิโรฒ] http://thesis.swu.ac.th/swuthesis/Mark/Ubsornsri_M.pdf

ฤทธิไกร ทวีเจริญ (2559). *ตัวแบบการพยากรณ์การลาออกของพนักงานฝ่ายผลิตในธุรกิจเครื่องฉีดพลาสติกของบริษัทแห่งหนึ่ง.* [การค้นคว้าอิสระปริญญามหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์] http://ethesisarchive.library.tu.ac.th/thesis/2016/TU_2016_5823036065_5595_6295.pdf

สื่ออิเล็กทรอนิกส์

กวิน เอี่ยมตระกูล. (2564). *การใช้ Google ประมาณค่าสถิติตลาดแรงงาน.* http://www.fpojournal.com/google-trend-forcast-labor-force-market/

ชิตพงษ์ กิตตินราดร. (2562). *Linear Regression Algorithm*. https://guopai.github.io/ml-blog02.html

ชิตพงษ์ กิตตินราดร. (2563). *Support Vector Machines*. https://guopai.github.io/ml-blog08.html

ชิตพงษ์ กิตตินราดร. (2563). *Neural Network Algorithm*. https://guopai.github.io/ml-blog14.html

ชิตพงษ์ กิตตินราดร. (2563). *Neural Network Programming*. https://guopai.github.io/ml-blog15.html

ณนนท์ นพรัตน์. (2563). *มาทำความรู้จักกับ การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning).* https://bigdata.go.th/big-data-101/introduction-to-reinforcement-

Ref. code: 25646304010116LCN

learning/#:~:text=Reinforcement-learning/#:~:text=Reinforcement%20Learning%20เป็น%20Machine%20Learning,Agent%20–%20ผู้กระทำ%20Action

ธนาคารแห่งประเทศไทย. (2564). *อัตราการว่างงานจำแนกตามภาค*. https://www.bot.or.th/Thai/Segmentation/Business/Pages/UnemploymentRate.aspx

ธนัตถ์พราว ลีลารัตน์, ณัฐพงษ์ พัฒนพงษ์ และ วศิน ศิวสฤษดิ์. (2564). *วิถีใหม่ในการพยากรณ์อัตราการว่างงานไทย ด้วยข้อมูลจาก Google Trends และ Twitter*. http://www.fpojournal.com/google-trends-twitter-forcasting/

มัทนา วิบุญศักดิ์. (2561). เมื่อเราพูดถึง AI คุณนึกถึงอะไร?. https://www.aware.co.th/พูดถึง-ai-คุณนึกถึงอะไร/

มัทนา วิบุญศักดิ์. (2563). Machine Learning คืออะไร?. https://www.aware.co.th/machine-learning-คืออะไร/#:~:text=Machine%20Learning%20คือ%20ส่วนการ,เครื่อง)%20เรียนรู้จากข้อมูล

ระบบเตือนภัยด้านแรงงาน กองเศรษฐกิจการแรงงาน สำนักงานปลัดกระทรวงแรงงาน. (2564). *ข้อมูลเศรษฐกิจแรงงาน ประจำเดือนตุลาคม 2564*. http://www.warning.mol.go.th/article/review

รุจเรขา วิทยาวุฒิกุล. (2554). *พฤติกรรมสารสนเทศ (Information Behavior)* https://ruchareka.wordpress.com/2011/07/15/พฤติกรรมสารสนเทศ-information-behaviour/#:~:text=T.D.%20Wilson%27s%20Information%20Behaviour%20model,เข้าใจ%20เพื่อจัดระเบียบ%20วางแผน

วรประภา นาควัชระ. (2559). *Nowcasting 101*. https://www.bangkokbiznews.com/blogs/columnist/113423

สำนักงานสถิติแห่งชาติ. (2564). *ภาวการณ์ทำงานของประชากร.* https://www.service.nso.go.th/nso/nsopublish/download/files/lfsSum64.

สำนักงานสถิติแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม. (2564). *การสำรวจภาวะการทำงานของประชากรทั่วราชอาณาจักร ไตรมาส 2 : เมษายน - มิถุนายน 2564.* http://www.nso.go.th/sites/2014/Pages/สำรวจ/ด้านสังคม/แรงงาน/ภาวะการทำงานของประชากร.aspx

Ref. code: 25646304010116LCN

อริศรา โชติวานิช. (2562). *Botnoi Classroom — มาทำให้ผลลัพธ์ดีขึ้นด้วย Ensemble Method กันเถอะ!*. https://medium.com/botnoi-classroom/botnoi-classroom-มาทำให้ผลลัพธ์ดีขึ้นด้วย-ensemble-method-กันเถอะ-b5ac9acfa8d3

Google Trends. (2564). สำรวจ Google Trends. https://trends.google.co.th/trends/explore

Phanpaporn Z. (2564). *ทำความรู้จัก “Linear Regression” Algorithm ที่คนทำ Machine Learning ยังไงก็ต้องได้ใช้!*. https://www.borntodev.com/2021/08/26/ทำความรู้จัก-linear-regression-algorithm-ที่คนท/

Thai programmer. (2561). *อะไรคือการเรียนรู้ของเครื่อง (Machine Learning)?*. https://www.thaiprogrammer.org/2018/12/อะไรคือ-การเรียนรู้ของเ/#:~:text=Machine%20Learning%20สามารถถูกแบ่ง,มีผู้สอน%20(Supervised%20learning)

TNN. (2564). *โควิดกระทบหนัก! สภาพัฒน์ห่วงคนตกงานมีแนวโน้มว่างงานนานขึ้น*. https://www.tnnthailand.com/news/wealth/89068/

Theses

Parkpoom tantaopas (2014). *ATTENTION EFFECT VIA INTERNET SEARCH INTENSITY IN ASIA-OCEANIA STOCK MARKETS.* [An independent study fulfillment of the requirements for the degree of master, Thammasat University] https://www.sciencedirect.com/science/article/pii/S0927538X16300385

Articles

Hyunyoung, C., & Varian, H. (2012). Predicting the Present with Google Trends. *The economic society of Australia*, 88, 2-9.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). An Introduction to Statistical Learning with Applications in R, 327-333.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396.

Ref. code: 25646304010116LCN

McLaren, N. & Rachana, S. (2011). Using Internet Search Data as Economic Indicators. *Bank of England Quarterly Bulletin*, No. 2011 Q2, 1-7.

Moreire, J. M., Soares, C., Jorge, A. M., & Sousa, J. F. (2012). Ensemble Approaches for Regression: a Survey. *ACM Journals*, 45(1), 10-14.

Nakavachara, V., & Lekfuangfu, N. (2018). Predicting the Present Revisited: The Case of Thailand. *Economics The Open-Access Open-Assessment E-Journal*, 12, 1-9.

Preis, T., Moat, H. S., & Stanlay, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *SCIENTIFIC REPORTS*, 2-6.

Puttanapong, N., Martinez Jr., A. M., Addawe, M., Bulan, J., Durante, R. L., & Martillan, M. (2020). Predicting Poverty Using Geospatial Data in Thailand. *ADB ECONOMICS WORKING PAPER SERIES,* 630, 17-18.

Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble Classification and Regression – Recent Developments, *Applications and Future Directions*, 11(1), 41-53.

Stephens-Davidowitz, S., & Varian, H. (2015). A Hands on Guide to Google Data. Google Inc, 2-23.

Tantaopas, P., Padungsaksawasdi, C., & Treepongkaruna, S. (2016). Attention effect via internet search intensity in Asia-Pacific stock markets. *Pacific Basin Finance Journal*, 38, 107-124.

Wang, Y., Wang, B., & Zhang, X. (2012). A New Application of the Support Vector Regression on the Construction of Financial Conditions Index to CPI Prediction. *Procedia Computer Science, 9*, 1263-1272.

Wilson, T. D. (2000). Human Information Behavior. *Special Issue on Information Science Research*, 3(2), 1–7.

Zhang, X.-r., Hu, L.-y., & Wang, Z.-s. (2010). Multiple kernel support vector regression for economic forecasting. *International Conference on Management Science & Engineering 17th Annual Conference Proceedings* (pp. 129-134). IEEE.

Ref. code: 25646304010116LCN

Electronic Media

CEIC. (2021). *Labour Force Employ and Unemployment*. https://insight.ceicdata.com/Untitled-insight/view

Dubey, B. (2020). *What is Support Vector Regression — SVR?.* Medium. https://medium.com/@bhartendudubey/what-is-support-vector-regression-svr-760b501b6141

Mc, C. (2017). *Machine learning fundamentals (II): Neural networks*. Towards Data Science. https://towardsdatascience.com/machine-learning-fundamentals-ii-neural-networks-f1e7b2cb3eef

Pedamkar, P. (2018). *Support Vector Regression.* EDUCBA. https://www.educba.com/support-vector-regression/

Sharp, T. (2020). *An Introduction to Support Vector Regression (SVR).* Towards Data Science. https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

Yuda. M. (2021). *Thailand's COVID crisis exposes the reality of '2% unemployment'.* NIKKEI Asia. https://asia.nikkei.com/Spotlight/Asia-Insight/Thailand-s-COVID-crisis-exposes-the-reality-of-2-unemployment

Ref. code: 25646304010116LCN