ระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง

พรทิวา วิศิษฏ์สรอรรถ

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2564

A WEB-BASED SYSTEM FOR TIME SERIES FORECAST

PORNTIVA VISITSORA-AT

An Independent Study Submitted in Partial Fulfillment of the

Requirements for the Degree of Master of Big Data Engineering,

College of Innovative Technology and Engineering,

Dhurakij Pundit University

Academic Year 2021

ใบรับรองงานสารนิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

หัวข้อสารนิพนธ์ ระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง

เสนอโดย พรทิวา วิศิษฏ์สรอรรถ

สาขาวิชา วิศวกรรมข้อมูลขนาดใหญ่

อาจารย์ที่ปรึกษาสารนิพนธ์ ดร.เอกสิทธิ์ พัชรวงศ์ศักดา

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

.....ประธานกรรมการ

(ดร.สรรพฤทธิ์ มฤคทัต)

..... กรรมการและอาจารย์ที่ปรึกษา

(ดร.เอกสิทธิ์ พัชรวงศ์ศักดา)

..... กรรมการ

(ดร.ธนภัทร มังคะจิตร)

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์รับรองแล้ว

.....

(ดร.ชัยพร เขมะภาคะพันธ์)

คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

วันที่ 30 เดือน พฤศจิกายน พ.ศ. 2564

หัวข้อสารนิพนธ์	ระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง
ชื่อผู้เขียน	พรทิวา วิศิษฏ์สรอรรถ
อาจารย์ที่ปรึกษา	ดร.เอกสิทธิ์ พัชรวงศ์ศักดา
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2564

บทคัดย่อ

การวิจัยนี้เป็นการศึกษามีวัตถุประสงค์เพื่อพัฒนาระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง โดยเปรียบเทียบวิธีพยากรณ์ออกเป็น 5 เทคนิคประกอบไปด้วยเทคนิค AutoRegressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average (SARIMA), วิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression), ต้นไม้ตัดสินใจ (Decision Tree) และ ป่าสุ่ม (Random Forest) ซึ่งพัฒนาขึ้นมาเป็นรูปแบบของเว็บแอปพลิเคชันที่ช่วยพยากรณ์ข้อมูลในอนาคตได้หลากหลายรูปแบบสำหรับงานวิจัยนี้ได้นำข้อมูลจากกรณีศึกษาต่าง ๆ มาประยุกต์ใช้สร้างแบบจำลองดังนี้ 1. ข้อมูลยอดขายรายวันของร้านกาแฟ 2. ข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม 3. ข้อมูลจำนวนผู้ป่วยและผู้เสียชีวิตรายใหม่จากสถานการณ์ COVID-19 ในประเทศไทย เมื่อแบ่งข้อมูลเพื่อสร้างแบบจำลองพยากรณ์ยอดขาย, จำนวนผู้ใช้บริการและจำนวนยอดผู้เสียชีวิตรายใหม่โดยใช้การคัดเลือกเทคนิคที่เหมาะสมที่สุดสำหรับการพยากรณ์โดยใช้เกณฑ์ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE) ต่ำที่สุดผลวิจัยพบว่าหากข้อมูลที่นำมาสร้างแบบจำลองมีอิทธิพลต่อฤดูกาลเทคนิค SARIMA ให้ผลการพยากรณ์ที่ดีกว่าเทคนิค ARIMA เช่น กรณีศึกษาการพยากรณ์จำนวนผู้ใช้บริการรถไฟฟ้ามหานครสายฉลองธรรมรายเดือนให้ค่า MAPE สำหรับเทคนิค SARIMA, ARIMA เท่ากับ 25.90% , 35.15% ตามลำดับ ส่วนการเปรียบเทียบเทคนิคระหว่าง Decision Tree, Random Forest และMultiple Linear Regression สำหรับการพยากรณ์ยอดขายรวมรายวันแยกตามรายสินค้า และจำนวนผู้เสียชีวิตรายใหม่จากสถานการณ์ COVID-19 ในประเทศไทยพบว่าเทคนิคให้ค่า MAPE เฉลี่ยต่ำที่สุดคือเทคนิค Random Forest 18.74% - 54.15% เมื่อเทียบกับเทคนิค Decision Tree 22.45% - 57.34% และ Multiple Linear Regression 28.08% - 156.86% ตามลำดับ

An Independent Study Title	A WEB-BASED SYSTEM FOR TIME SERIES FORECAST
Author	Porntiva Visitsora-at
An Independent Study Advisor	Dr. Eakasit Pacharawongsakda
Department	Big Data Engineering
Academic Year	2021

ABSTRACT

This research was conducted to develop a web-based system for time series analysis using statistic and machine learning techniques. The five compared models of time series analysis included AutoRegressive Integrated Moving Average (ARIMA) and Seasonal AutoRegressive Integrated Moving Average (SARIMA), Multiple Linear Regression, Decision Tree, and Random Forest. The system was developed as a web application which could forecast future information in various datasets. In this study, three datasets were utilized for model development: 1. data on daily sales of a coffee shop, 2. data on passengers of the M.R.T Chalong Ratchadham Line, and 3. data on new cases and deaths of COVID-19 in Thailand. Having separated data for modeling sales forecast, the number of passengers, and the number of new COVID-19 deaths in Thailand, through selection of the most appropriate forecasting techniques based on the lowest Mean Absolute Percentage Error (MAPE), the results showed that when the data used for model development influenced SARIMA, it yielded better forecast in comparison to ARIMA. For example, a study forecasting the number of passengers of the M.R.T Chalong Ratchadham Line each month had MAPE of SARIMA, ARIMA representing 25.90%, 35.15% respectively. Moreover, the comparison between Decision Tree, Random Forest, and Multiple Linear Regression for forecasting the daily sales by product and new cases and deaths of COVID-19 in Thailand indicated that Random Forest had the lowest average of MAPE which was 18.74% - 54.15% compared to Decision Tree which was 22.45% - 57.34% and Multiple Linear Regression which was 28.08% - 156.86%, respectively.

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จสมบูรณ์ลุล่วงไปได้ด้วยดีเพราะได้รับความกรุณาชี้แนะและช่วยเหลืออย่างดียิ่งจาก ดร.เอกสิทธิ์ พัชรวงศ์ศักดาซึ่งเป็นอาจารย์ที่ปรึกษาสารนิพนธ์ ที่ได้กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขข้อบกพร่องต่าง ๆ มาโดยตลอดเพื่อให้สารนิพนธ์ฉบับนี้สมบูรณ์ ผู้เขียนจึงขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ผู้เขียนขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.สรรพฤทธิ์ มฤคทัต ที่กรุณาให้เกียรติเป็นประธานโดยมี ดร.ธนภัทร ฆังคะจิตรเป็นกรรมการในการสอบสารนิพนธ์ ซึ่งได้กรุณาตรวจแก้ไขสารนิพนธ์ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้นและ นางสาวกุลธิดา รอดบุญ รวมถึงเจ้าหน้าที่บัณฑิตมหาวิทยาลัยธุรกิจบัณฑิตย์ ทุกท่านที่ให้ความสะดวกด้านอำนวยการ และประสานงาน ในการทำสารนิพนธ์ให้กับผู้เขียน ทำให้การจัดทำสารนิพนธ์ของผู้เขียนในครั้งนี้สำเร็จลุล่วงไปด้วยดี

ผู้เขียนขอขอบพระคุณนางสาวพรทิพย์ วิศิษฎ์สรอรรถ ที่กรุณาให้เก็บรวบรวมข้อมูลของร้านกาแฟเพื่อนำมาใช้ทดลองจริงจนทำให้สารนิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี

ขอขอบพระคุณบิดามารดาที่สนับสนุนและให้กำลังใจจนงานสารนิพนธ์สำเร็จด้วยดี ผู้เขียนขอน้อมบูชาพระคุณบิดามารดาและบูรพาจารย์ทุกท่านที่ได้อบรมสั่งสอนวิชาความรู้ และให้ความเมตตาผู้เขียนมาโดยตลอดและเป็นเป็นกำลังใจสำคัญที่ทำให้สารนิพนธ์สำเร็จลุล่วงไปด้วยดี

ขอขอบคุณเพื่อน ๆ BD4 ทุกท่านที่คอยเป็นกำลังใจและช่วยเหลือจนงานสารนิพนธ์สำเร็จด้วยดี

พรทิวา วิศิษฎ์สรอรรถ

สารบัญ

สารบัญ (ต่อ)

สารบัญตาราง

ตารางที่ **หน้า**

สารบัญภาพ

ภาพที่ **หน้า**

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

แนวคิดอนุกรมเวลาปรากฏเป็นพื้นฐานในแทบทุกกิจกรรมมีการประยุกต์ใช้การวิเคราะห์อนุกรมเวลาในทางปฏิบัติหลายอย่าง เช่น การคาดการณ์ทางเศรษฐกิจ การวิเคราะห์ตลาดหุ้น การพยากรณ์แผ่นดินไหว หรือการคาดการณ์ผลตอบแทนและอื่นๆ นอกจากนี้การวิเคราะห์อนุกรมเวลาช่วยให้เราเข้าใจว่าอะไรคือแรงผลักดันสำคัญที่นำไปสู่แนวโน้มเฉพาะในจุดข้อมูลอนุกรมเวลา

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง โดยแบ่งเป็น 1. Autoregressive Integrated Moving Average (ARIMA) 2. Seasonal Autoregressive Integrated Moving Average (SARIMA) ประเภทแบบจำลองอนุกรมเวลา เพื่อจัดการกับข้อมูลที่มีลักษณะเกี่ยวข้องกันทางด้านเวลานั่นคือข้อมูลในอดีตสามารถส่งผลต่อข้อมูลในอนาคตได้ 3. เทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) เป็นการวิเคราะห์ความสัมพันธ์ในรูปแบบเชิงเส้นตรง (Linearity) ระหว่างตัวแปรอิสระ (Independent Variable) ที่มีมากกว่า 1 ตัวและตัวแปรตาม (Dependent Variable) 4. เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และ 5. เทคนิคป่าสุ่ม (Random Forest) มาพยากรณ์ข้อมูล ซึ่งหลักๆที่เลือกใช้หลายเทคนิคจุดประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคว่าเทคนิคใดให้ความคาดเคลื่อนในการพยากรณ์ต่ำที่สุดเพื่อนำไปใช้จริงสำหรับข้อมูลชุดใหม่ๆที่ต้องการพยากรณ์ได้

ผลที่ได้จากงานวิจัยนี้หลังจากที่นำเข้าอนุกรมข้อมูลหลากหลายประเภท เช่น ข้อมูลยอดขาย ข้อมูลจำนวนผู้ใช้บริการรถไฟฟ้า รวมถึงข้อมูลยอดผู้ติดเชื้อ COVID-19 เข้าสู่ระบบเพื่อทำการประมวลผลสร้างแบบจำลองด้วยเทคนิคต่างๆข้างต้นจะแสดงผลเปรียบเทียบค่าความคาดเคลื่อนทำให้ผู้ใช้งานสามารถเลือกใช้แบบจำลองในการพยากรณ์ข้อมูลในอนาคตที่มีความแม่นยำ งานวิจัยนี้แบ่งข้อมูลเพื่อสร้างแบบจำลองที่แตกกันคือ

แบบที่ 1 คือ ชุดข้อมูลอนุกรมเวลาที่มีตัวแปรสร้างแบบจำลอง คือ เวลาและค่าที่ต้องการพยากรณ์เพื่อใช้กับเทคนิค ARIMA, SARIMA

แบบที่ 2 คือ ชุดข้อมูลอนุกรมเวลาที่มีตัวแปรสร้างแบบจำลอง คือ เวลา ตัวแปรอิสระร่วม และค่าที่ต้องการพยากรณ์เพื่อใช้กับเทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) ต้นไม้ตัดสินใจ (Decision Tree) และ ป่าสุ่ม (Random Forest)

ใช้ตัวชี้วัดการประเมินผลด้วยค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE)

1.2 วัตถุประสงค์ของการวิจัย

พัฒนาเว็บแอปพลิเคชันเพื่อนำเสนอรูปแบบเปรียบเทียบการพยากรณ์ข้อมูลด้วยเทคนิค Autoregressive Integrated Moving Average (ARIMA) Seasonal Autoregressive Integrated Moving Average (SARIMA) เทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และวิธีป่าสุ่ม (Random Forest) โดยมุ่งเน้นให้แบบจำลองจากเทคนิคต่างๆข้างต้นสามารถวิเคราะห์คุณลักษณะสำคัญจากอนุกรมข้อมูลที่นำมาใช้และพิจารณาความสัมพันธ์ของแต่ละคุณลักษณะของข้อมูลในอดีตมาเพิ่มประสิทธิภาพการพยากรณ์ให้มีความแม่นยำและความคาดเคลื่อนที่ต่ำที่สุดเพื่อนำไปใช้ในการพยากรณ์ในอนาคตที่มีความน่าเชื่อถือ

1.3 ขอบเขตการวิจัย

1.3.1 ข้อมูลกรณีศึกษาที่นำมาใช้ในงานวิจัยคือ

1.3.1.1 ข้อมูลยอดขายจริงรายวันของร้านกาแฟตั้งแต่เดือน มีนาคม ถึง สิงหาคม ปี พ.ศ. 2564 รวมไปถึงข้อมูลเพิ่มเติมดังนี้

1.3.1.1.1 ข้อมูลอุณหภูมิ[1] ตั้งแต่เดือน มีนาคม ถึง สิงหาคม ปี พ.ศ. 2564

1.3.1.1.2 ข้อมูลวันหยุดตามประเพณีของสถาบันการเงิน[2] ปี พ.ศ. 2564

---

[1]Bangkok, Bangkok, Thailand Weather History,
จาก https://www.wunderground.com/history/monthly/th/bangkok

[2]วันหยุดตามประเพณีของสถาบันการเงิน ประจำปี,
จาก https://www.bot.or.th/Thai/FinancialInstitutions/FIholiday/Pages/2021.aspx

1.3.1.2 ข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร[3] สายฉลองรัชธรรมตั้งแต่เดือน สิงหาคม ปี 2559 ถึง สิงหาคม ปี พ.ศ. 2564

1.3.1.3 ข้อมูลจำนวนผู้ป่วย COVID-19 ระลอก 3[4] ตั้งแต่เดือน มิถุนายน ถึง ตุลาคม ปี พ.ศ. 2564

1.3.2. ทำการเตรียมข้อมูลให้สมบูรณ์ให้พร้อมสำหรับการนำเข้าไปใช้กับแบบจำลองต่าง ๆ

1.3.3. พัฒนาเว็บแอปพลิเคชัน

1.3.3.1 สร้างแบบจำลองพยากรณ์ข้อมูลด้วยการเปรียบเทียบด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่องในรูปแบบต่างๆ

1.3.3.2 ทำการวัดประสิทธิภาพของแต่ละแบบจำลองโดยพิจารณาจากค่าความค่าเคลื่อนด้วยค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE)

1.3.3.3 สร้างตัวเลือกการพยากรณ์ข้อมูลจากการเลือกแบบจำลองที่พิจารณาประสิทธิภาพเปรียบเทียบก่อนหน้าสำหรับข้อมูลใหม่

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 สามารถนำระบบพยากรณ์ข้อมูลที่มีประสิทธิภาพแม่นยำน่าเชื่อถือไปใช้ให้เกิดประสิทธิภาพในการวางแผนด้านการตลาด การจัดสรรทรัพยากร คาดการณ์เศรษฐกิจ หรือ อื่นๆ ตามจุดประสงค์ของข้อมูลในการใช้งานได้

1.4.2 ระบบพยากรณ์ข้อมูลมีหลากหลายเทคนิคเปรียบเทียบสำหรับสร้างแบบจำลองทำให้ผู้ใช้งานสามารถเลือกใช้เทคนิคการพยากรณ์ให้เหมาะสมกับจุดประสงค์การใช้งานของตนเองได้

1.4.3 ระบบใช้งานง่ายไม่ซับซ้อนและไม่จำเป็นต้องมีความรู้เกี่ยวกับเทคนิคระดับเชี่ยวชาญ ก็สามารถนำระบบนี้ไปใช้งานให้เกิดประโยชน์ได้

---

[3]ข้อมูลผู้โดยสาร โครงการรถไฟฟ้ามหานคร สายเฉลิมรัชมงคล และสายฉลองรัชธรรม จาก https://data.go.th/dataset/mrta-crmk

[4]รายงานสถานการณ์ COVID-19 ระลอก 3 (ตั้งแต่ 01/04/2021 –ปัจจุบัน) จาก https://covid19.ddc.moph.go.th/api/Cases/timeline-cases-all

บทที่ 2
แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

การศึกษาทฤษฎีที่เกี่ยวข้องกับงานวิจัยนี้ได้ศึกษาแนวคิด ทฤษฎีและผลงานวิจัยที่เกี่ยวข้องเพื่อนำมาประยุกต์ใช้เพื่อเป็นความรู้พื้นฐานและใช้กำหนดแนวทางในการศึกษาวิจัย ประกอบไปด้วย 6 หัวข้อหลัก ดังนี้

2.1 ความหมาย ประโยชน์และความสำคัญของการพยากรณ์

2.1.1 ความหมายของการพยากรณ์

การคาดการณ์หรือทำนายสิ่งที่จะเกิดขึ้นในอนาคตโดยอาศัยการวิเคราะห์ข้อมูลในอดีตหรือข้อมูลปัจจุบันสามารถนำไปใช้เพื่อให้ทราบถึงแนวโน้มของการเปลี่ยนแปลงทิศทางการธุรกิจเศรษฐกิจหรือสภาพแวดล้อมที่จะมีผลกระทบได้ในอนาคตทำให้สามารถที่จะวางแผนหรือกำหนดนโยบายเพื่อให้บรรลุวัตถุประสงค์ของการดำเนินการได้

2.1.2 ประโยชน์และความสำคัญของการพยากรณ์

1. ลดโอกาสที่จะเกิดความเสี่ยงหรือไม่แน่นอนที่อาจเกิดขึ้นในอนาคต
2. ช่วยในการวางแผนการดำเนินงานรวมทั้งกำหนดเป้าหมาย วัตถุประสงค์ วิสัยทัศน์ แผนกลยุทธ์ และแผนปฏิบัติการได้อย่างเหมาะสม
3. ทำให้สามารถสรรหาทรัพยากรอื่นๆ มาเพิ่มเติมจากพื้นฐานข้อมูลที่มีอยู่ปัจจุบัน

2.2 องค์ประกอบ ประเภทและคุณลักษณะของการพยากรณ์

2.2.1 องค์ประกอบของการพยากรณ์ที่ดี (หทัยชนก นานานอก, 2553)

1. ระบุวัตถุประสงค์ในการนำผลการพยากรณ์ไปใช้และกำหนดช่วงเวลาที่การพยากรณ์จะครอบคลุมถึงเพื่อที่จะสามารถเลือกใช้วิธีในการพยากรณ์ได้อย่างถูกต้องเหมาะสม
2. รวบรวมข้อมูลอย่างมีระบบถูกต้องตามความเป็นจริง เพราะคุณภาพของข้อมูลมีผลสำคัญมากต่อการพยากรณ์
3. เมื่อมีสินค้าหลายชนิดควรจำแนกประเภทของสินค้าที่มีลักษณะตามความต้องการที่คล้ายคลึงกันไว้เป็นกลุ่มเดียวกันโดยแบ่งเป็นพยากรณ์สำหรับกลุ่มแล้วจึงแยกพยากรณ์สำหรับแต่ละสินค้าในกลุ่มอีกครั้งโดยเลือกวิธีการพยากรณ์ที่เหมาะสมกับแต่ละกลุ่มและสินค้า

4. ควรระบุข้อจำกัดและสมมติฐานที่ตั้งไว้ในการพยากรณ์นั้นๆ เพื่อผู้นำผลการพยากรณ์ไปใช้จะได้ทราบถึงเงื่อนไขข้อจำกัดที่มีผลต่อการวิเคราะห์ผลของค่าพยากรณ์

5. หมั่นตรวจสอบความถูกต้องแม่นยำของค่าพยากรณ์ได้กับค่าจริงที่เกิดขึ้นเป็นอย่างสม่ำเสมอเพื่อปรับวิธีการที่ใช้ในการคำนวณให้เหมาะสม

2.2.2 ประเภทและคุณลักษณะของการพยากรณ์

แบ่งตามระยะเวลาที่ใช้ในการพยากรณ์

1. การพยากรณ์ระยะสั้น (Short-Range Forecasting)

ระยะเวลาปกติน้อยกว่า 1 ฤดูกาล 1 วัน ไปถึง 1 ปีลักษณะเป็นการพยากรณ์ลงลึกถึงระดับเฉพาะเพื่อการวางแผนกิจกรรมและการปรับปรุงการจัดซื้อและการจัดการสินค้าคงเหลือเหมาะนำไปใช้สำหรับการควบคุมระยะสั้นวิธีพยากรณ์ส่วนใหญ่จะใช้การคาดการณ์แนวโน้มด้วยกราฟการปรับเรียบแบบเอ็กซ์โพเนนเชียลการใช้ดุลยพินิจ

2. การพยากรณ์ระยะปานกลาง (Medium-Range Forecasting)

ระยะเวลาปกติเป็นฤดูกาลจนไปถึง 3 ปี ลักษณะเป็นการพยากรณ์เป็นเชิงตัวเลขมักต้องการความถูกต้องน่าเชื่อถือเหมาะนำไปใช้สำหรับการวางแผนโดยรวมวิธีพยากรณ์ส่วนใหญ่คือการรวบรวมข้อคิดเห็นอนุกรมเวลาการวิเคราะห์การถดถอย ดัชนีทางเศรษฐกิจ การใช้ดุลยพินิจ

3. การพยากรณ์ระยะยาว (Long-Range Forecasting)

ระยะเวลาปกติ 5 ปีหรือมากกว่า 3 ปีขึ้นไป ลักษณะเป็นการพยากรณ์กว้างๆและมักเป็นเชิงคุณภาพเหมาะนำไปใช้สำหรับการวางแผนธุรกิจวิธีพยากรณ์ส่วนใหญ่จะนำไปใช้เกี่ยวกับด้านเทคโนโลยีภาวะเศรษฐกิจการศึกษาตลาดการใช้ดุลพินิจ

2.2.3 วิธีการพยากรณ์

2.2.5.1 การพยากรณ์เชิงคุณภาพ (Qualitative Methods หรือ Objective Methods)

เป็นกลุ่มของวิธีการพยากรณ์ที่นำข้อมูลและวิธีการเชิงคุณภาพ พยากรณ์โดยผู้ที่มีประสบการณ์ความรู้ความสามารถโดยไม่ใช้ตัวแบบทางคณิตศาสตร์จึงตรวจสอบความแม่นยำของการพยากรณ์ได้ยากกว่าการพยากรณ์เชิงปริมาณตัวอย่างเทคนิคที่ใช้ในการพยากรณ์เชิงคุณภาพ เช่น การประมาณการ การระดมความคิด การสำรวจตลาด หรือเทคนิคเดลไฟ (Delphi) คือสามารถแสดงความคิดเห็นแบบไม่แสดงตัวตนเมื่อได้คำตอบและได้ข้อสรุปไม่ตรงกันก็จะทำใหม่เรื่อย ๆ จนกว่าได้ข้อสรุปที่เป็นหนึ่งเดียว

2.2.5.2 การพยากรณ์เชิงปริมาณ (Quantitative Methods)

เป็นกลุ่มของวิธีการพยากรณ์ที่นำข้อมูล หรือตัวเลขจากอดีตโดยใช้รูปแบบทางคณิตศาสตร์ เพื่อใช้ในการสร้างแบบจำลองและพยากรณ์ไปในอนาคตโดยแบ่งตัวอย่างออกเป็น 2 เทคนิคย่อย

1. การพยากรณ์ความสัมพันธ์ (Casual Forecasting)

เป็นเทคนิคที่ใช้ปัจจัยที่คาดว่าจะมีความสัมพันธ์กับตัวแปรที่จะพยากรณ์ เช่น การพยากรณ์ยอดขายจะพิจารณาหาความสัมพันธ์ระหว่างยอดขายกับค่าโฆษณาจำนวนช่องทางการจัดจำหน่ายรายได้เฉลี่ยของประชากรคุณภาพของสินค้า เป็นต้น การหาความสัมพันธ์ดังกล่าวจะใช้เทคนิคที่เรียกว่าการวิเคราะห์ความถดถอยและสหสัมพันธ์

2. การพยากรณ์อนุกรมเวลา (Time series Forecasting)

เป็นเทคนิคที่ใช้ข้อมูลในอดีตของตัวแปรที่ต้องการพยากรณ์เพื่อพยากรณ์ค่าในอนาคต เช่น ใช้ข้อมูลจำนวนผู้เข้าใช้บริการของปี พ.ศ. 2559-2564 เพื่อพยากรณ์จำนวนผู้เข้าใช้บริการของปี พ.ศ. 2565

2.3 เทคนิคการวิเคราะห์การถดถอย

2.3.1 การวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression Analysis)

เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไปกับผลของตัวแปรตามที่เกิดขึ้นโดยมีรูปแบบสมการพยากรณ์ถดถอยพหุคูณดังนี้

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + + b_k x_k$$

โดยที่สัญลักษณ์ที่ใช้มีความหมายดังนี้

x_i คือ ค่าของตัวแปรอิสระ

$\hat{y}$ คือ ค่าของตัวแปรตาม

b_0 คือ ค่าคงที่ (Constant) ของสมการถดถอยโดยจะเป็นจุดตัดแกน Y ของสมการ

b_i คือ ค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรอิสระ x_i แต่ละตัว

k คือ จำนวนของตัวแปรอิสระ

(ดร.สุทิน, 2560) กล่าวว่า ในการวิเคราะห์การถดถอยจำเป็นต้องมีข้อตกลง (Assumption) ในการวิเคราะห์การถดถอย ซึ่งมีจำนวนข้อตกลงเบื้องต้นที่สำคัญ เช่น

1. ตัวแปรอิสระและตัวแปรตามต้องเป็นตัวแปรเชิงปริมาณ (Quantitative Variable) หรือ ตัวแปรต่อเนื่อง (Continuous Variable) หรือมีระดับการวัดเป็น Interval หรือ Ratio Scale เช่น ส่วนสูงน้ำหนัก เป็นต้น ในกรณีที่ตัวแปรอิสระมีระดับการวัดเป็น Nominal หรือ Ordinal Scale

จะต้องแปลงข้อมูลให้เป็นตัวเลขคือ มีค่า 0 กับ 1 ก่อนจึงจะสามารถนำไปวิเคราะห์แต่ไม่ควรจะมีการแปลงข้อมูลในรูปแบบดังกล่าวหลายตัวเพราะจะทำให้เกิดความคลาดเคลื่อนในการพยากรณ์ได้มากขึ้น

2. ตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์เชิงเส้นตรง

3. ตัวแปรอิสระควรเป็นอิสระต่อกัน (ค่าสหสัมพันธ์ไม่ควรเกิน 0.7) เพราะจะทำให้เกิด Multicollinearity คือ การที่ตัวแปรอิสระบางตัวมีความสัมพันธ์กันเองซึ่งส่งผลกระทบให้ค่าสัมประสิทธิ์การตัดสินใจ (R^2) สูงเกินกว่าความเป็นจริง

4. มีตัวแปรตามมีการแจกแจงแบบปกติ (Normal Distribution)

5. ค่าความแปรปรวนของตัวแปรตามที่คงที่

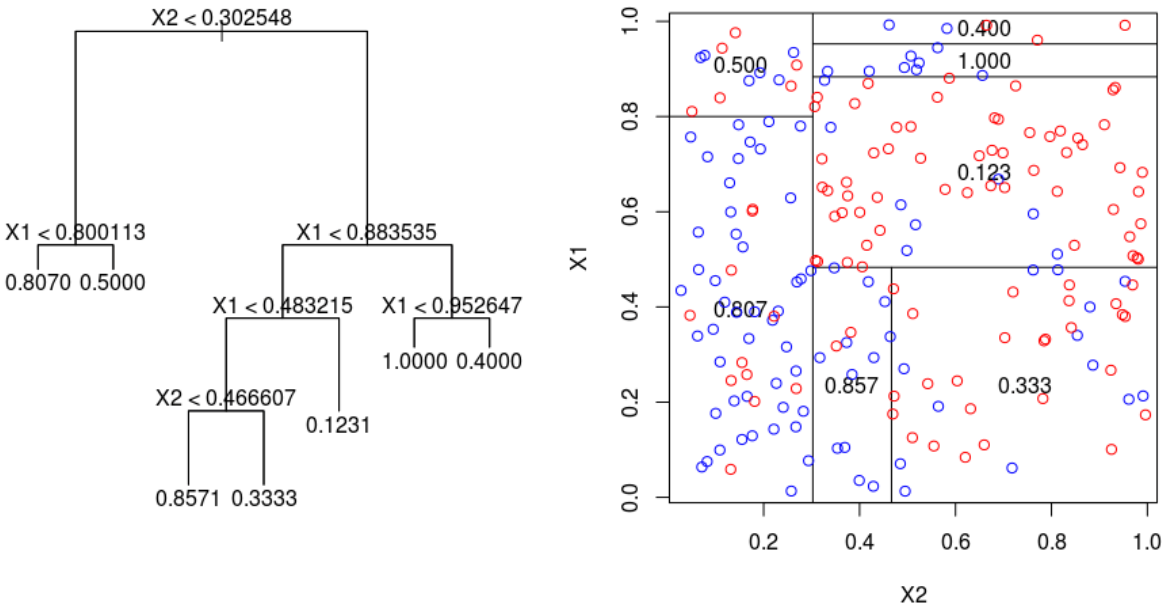
6. ความแปรปรวนของค่าความคลาดเคลื่อนจากการพยากรณ์ (Residual) ทุกจุดบนเส้นถดถอยมีค่าเท่ากัน

2.3.2 เทคนิคต้นไม้ตัดสินใจและป่าสุ่ม รูปแบบการถดถอย (Decision Tree and Random Forest Regression)

2.3.2.1 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning) รูปแบบหนึ่งที่ได้รับความนิยมโดยเทคนิคการให้ผลลัพธ์จะออกมาในรูปแบบลักษณะของโครงสร้างต้นไม้เป้าหมายคือการสร้างแบบจำลองที่คาดการณ์ค่าของตัวแปรตามโดยการเรียนรู้กฎการตัดสินใจอย่างง่ายที่อนุมานจากคุณลักษณะข้อมูลโดยจำลองแผนผังการตัดสินใจโดยเลือกแอตทริบิวต์ที่มีอิทธิพลต่อการจำแนกข้อมูลสูงที่สุด (Root node) ไปยังโหนดถัดไปจนกว่าจะไม่มีเงื่อนไขให้ไปต่อนั่นถือถึงใบ (Leaf node)

อัลกอริทึมที่ใช้สร้างแบบจำลองแผนผังต้นไม้ตัดสินใจมีดังนี้ ID3, C4.5, C5.0 และ CART โดยงานวิจัยนี้จะใช้ไลบรารี่ของ scikit-learn ที่ใช้ CART (Classification and Regression Trees) (Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J.) (1984) ซึ่งสนับสนุนการพยากรณ์ในรูปแบบการถดถอย (Regression) อัลกอริทึม CART (Classification and Regression Trees) มีลักษณะของต้นไม้ตัดสินใจในรูปแบบของต้นไม้ไบนารี่ (Binary tree) ซึ่งแต่ละโหนดจะแตกออกเป็น 2 กิ่งเท่านั้นดังภาพที่ 2.1 โดยคำนวณดัชนีจินี (Gini Index) สำหรับเลือกแอตทริบิวต์ที่จะเป็นโหนดราก (Root node) และใบ (Leaf node)



ภาพที่ 2.1 ลักษณะของต้นไม้ตัดสินใจในรูปแบบของต้นไม้ไบนารี่ (Binary tree)

ที่มา: https://gdcoder.com/decision-tree-regressor-explained-in-depth/

2.3.2.2 ป่าสุ่ม (Random Forest)

(ธนัท, 2561) ป่าสุ่มเป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning) รูปแบบหนึ่งที่ใช้พื้นฐานพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) ในทำนายในรูปแบบชุดของต้นไม้ตัดสินใจหลายๆ ต้น (Ensemble of Decision Trees) โดยสร้างจากการสุ่มข้อมูลตัวอย่างแบบเลือกแล้วใส่กลับ (random sampling with replacement) เพื่อนำมาสร้างเป็นแบบจำลองต้นไม้โดยแต่ละต้นมีลักษณะที่ไม่ซ้ำกันโดยแต่ละแบบจะจำลองจะมีการทำนายผลซึ่งผลจากการทำนายของต้นไม้แต่ละต้นจะทำการโหวตเลือกผลการทำนายที่ได้รับการโหวตมากที่สุดวิธีการนี้เรียกว่า Bagging หรือ Bootstrapping โดยงานวิจัยนี้ใช้ Random forest ของ scikit-learn เช่นเดียวกันกับเทคนิคต้นไม้ตัดสินใจ

2.4 การพยากรณ์ในรูปแบบอนุกรมเวลา

การวิเคราะห์อนุกรมเวลา (Time series analysis) ซึ่งเป็นวิธีการอาศัยข้อมูลจากอดีตเพื่อการพยากรณ์หรือคาดหมายสิ่งที่เกิดขึ้นในอนาคต

องค์ประกอบของอนุกรมเวลา (Time Series Components) ประกอบด้วย 4 ส่วนดังนี้

1. แนวโน้ม (Trend Component) มีรูปแบบการเพิ่มขึ้นหรือลดลงอย่างต่อเนื่องจะแสดงรูปแบบให้เห็นได้เมื่อมีระยะเวลาหลาย ๆ ปี

2. วัฏจักร (Cyclical Component) มีรูปแบบการเคลื่อนไหวซ้ำ ๆ ของการเพิ่มขึ้นหรือลดลงจะแสดงรูปแบบให้เห็นได้เมื่อมีระยะเวลามากกว่า 1 ปีอาจจะเกิดจากผลกระทบจากองค์ประกอบทางเศรษฐกิจ หรือวัฏจักรของธุรกิจ เป็นต้น

3. ฤดูกาล (Seasonal Component) มีรูปแบบความผันผวนขึ้นลงที่เกิดขึ้นเป็นประจำมีการเปลี่ยนแปลงตามฤดูกาล เป็นต้นจะแสดงรูปแบบให้เห็นเมื่อมีระยะเวลาประมาณ 1 ปี

4. การสุ่ม/ผิดปกติ (Random/Irregular Component) มีรูปแบบที่ไม่มีแบบแผนเป็นผลกระทบจากการเปลี่ยนแปลงแบบสุ่มหรือเหตุการณ์ที่ไม่คาดคิดมาก่อนเกิดขึ้นในช่วงเวลาสั้น ๆ และอาจจะไม่เกิดขึ้นซ้ำอีก

2.4.1 การพยากรณ์ด้วยเทคนิค Box-Jenkins

2.4.1.1 Autoregressive integrated moving average (ARIMA) เป็นเทคนิคพยากรณ์วิเคราะห์ข้อมูลอนุกรมเวลาอาศัยพฤติกรรมของข้อมูลในอดีตเพื่อกำหนดรูปแบบในปัจจุบันและอธิบายแนวโน้มหรือปรากฏการณ์ต่างๆ ของตัวข้อมูลเองในอนาคตโดยการหารูปแบบที่เหมาะสมให้กับข้อมูลเพื่อใช้กับการพยากรณ์โดยจะใช้เกณฑ์ AIC เลือกแบบจำลองที่ให้ค่า AIC น้อยที่สุดเป็นแบบจำลองที่ใช้ของข้อมูลอนุกรมเวลาที่พิจารณา

(ผศ.ดร. เฉลิมพล,2562) แบบจำลอง ARIMA ประกอบด้วยองค์ประกอบ 3 ส่วน คือ

1. AR(p) คือ Y_t ที่ถูกกำหนดด้วย Autoregressive process

   โดย $Y_t = Y_{t-1}, Y_{t-2},\ldots, Y_p$

2. I(d) คือ อันดับความหยุดนิ่งของ Y_t

   โดย $d = 0, 1, 2\ldots, d$

3. MA(q) คือ Y_t ที่ถูกกำหนดด้วย Moving average process

   โดย $Y_t = \varepsilon_{t-1}, \varepsilon_{t-2},\ldots, \varepsilon_q$

p และ q หมายถึง ลำดับของคาบเวลาในอดีต (Lag length) ที่เหมาะสม

Autoregressive process หรือ AR(p) เป็นกระบวนการของค่าสังเกตค่าหนึ่ง Y_t ที่ถูกกำหนดขึ้นจากความสัมพันธ์ของตัวมันเองในอดีต

$$Y_t = \alpha_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t$$

Y คือ ตัวแปรตาม, α คือ ค่าคงที่, φ คือ ค่าสัมประสิทธิ์ของ AR, t คือ เวลา, p คือ ลำดับของคาบเวลาในอดีต

Moving Average process หรือ MA(q) หมายถึง รูปแบบที่แสดงว่าค่า Y_t ถูกกำหนดมาจากค่าความคลาดเคลื่อน (ε) ในอดีต

$$Y_t = \alpha_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

Y คือ ตัวแปรตาม, α คือ ค่าคงที่, θ คือ ค่าสัมประสิทธิ์ของ MA, t คือ เวลา, q คือ ลำดับของคาบเวลาในอดีต

Autoregressive and Moving Average Process หรือ ARMA(p,q) เป็นการรวมกันระหว่าง AR กับ MA นั่นคือข้อมูลอนุกรมเวลาขึ้นอยู่กับทั้งค่าของข้อมูลอนุกรมเวลาในอดีตและค่าความคลาดเคลื่อนทั้งในปัจจุบันและในอดีตเป็นวิธีวิเคราะห์อนุกรมเวลาที่อาศัยขบวนการ Stochastic โดยถือว่าข้อมูลที่เกิดขึ้นตามเวลาที่เปลี่ยนแปลงไปมีลักษณะการเกิดที่เป็นไปตามกฎความน่าจะเป็นซึ่งการวิเคราะห์อนุกรมเวลาโดยวิธีนี้ลักษณะของอนุกรมเวลาต้องเป็นอนุกรมเวลาที่มีคุณสมบัติ Stationary

แบบจำลอง ARMA เมื่อตัวแปร Y มีความหยุดนิ่ง ณ ระดับปกติของข้อมูล หรือ I(0), d = 0

ARMA(p,d) คือ $Y_t = \alpha_0 + \varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

แบบจำลอง ARIMA เมื่อ ตัวแปร Y มีความหยุดนิ่ง ณ ผลต่างลำดับที่ d หรือ I(d)

ARIMA(p,d,q) คือ $\Delta^d Y_t = \alpha_0 + \varphi_1 \Delta^d Y_{t-1} + \ldots + \varphi_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

2.4.1.2 Seasonal Autoregressive integrated moving average (SARIMA) เป็นเทคนิคพยากรณ์วิเคราะห์ข้อมูลอนุกรมเวลาถูกพัฒนาขึ้นเพื่อรองรับการทำงานที่คำนึงถึงความผันแปรตามฤดูกาลซึ่งเป็นส่วนประกอบที่มีความสำคัญดังนั้นรูปแบบจะแสดงเป็น SARIMA(p,d,q)x(P,D,Q) และขั้นตอนการสร้างแบบจำลองพยากรณ์แสดงรายละเอียดดังนี้ (วรางคณา กีรติวิบูลย์ และเจ๊ะอัฐฟาน มาหิเละ, 2556)

$$\varphi_p(B)\Phi_p(B^s)(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

เมื่อ$\delta = \mu\varphi_p(B)\Phi_p(B^s)$ แทนค่าคงตัว โดยที่ μ คือค่าเฉลี่ยของอนุกรมเวลาที่คงที่ (Stationary)

$$\varphi_p(B) = 1 - \varphi_2 B^2 - \ldots - \varphi_p B^p$$

คือตัวดำเนินการสหสัมพันธ์ในตัวอันดับที่ p กรณีไม่มีฤดูกาล (Non-Seasonal Moving Average Operator of Order q): AR(p))

$$\Phi_p(B^s) = 1 - \Phi_2 B^{2s} - \ldots - \Phi_p B^{ps}$$

แทนตัวดำเนินการสหสัมพันธ์ในตัวอันดับที่ P กรณีมีฤดูกาล (Seasonal Moving Average Operator of Order Q): SAR(P))

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^q$$

แทนตัวดำเนินการเฉลี่ยเคลื่อนที่อันดับที่ q กรณีไม่มีฤดูกาล (Non-Seasonal Moving Average Operator of Order q): MA(q)

$$\Theta_p (B^s) = 1 - \Theta_1 B^s - \Theta_2 B^2 - \ldots - \Theta_Q B^{Qs}$$

แทนตัวดำเนินการเฉลี่ยเคลื่อนที่อันดับที่ Q กรณีมีฤดูกาล)Seasonal Moving Average Operator of Order Q): SMA(Q)

d และ D แทนลำดับที่ของการหาผลต่างและผลต่างฤดูกาลตามลำดับ

s แทนจำนวนคาบของฤดูกาล

B แทนตัวดำเนินการถอยหลัง)Backward Operator) โดยที่ $B^s Y_1 = Y_{1-s}$

ε_t แทนอนุกรมเวลาของความคลาดเคลื่อนที่มีการแจกแจงปกติและเป็นอิสระกัน

2.4.1.3 การเลือกค่า p,d,q และ P,D,Q ให้เหมาะสมกับแบบจำลอง

หาค่า d ด้วยวิธีการตรวจสอบอันดับความหยุดนิ่ง Augmented Dickey-Fuller (ADF) เพื่อที่จะหลีกเลี่ยงข้อมูลที่มีค่าเฉลี่ย (mean) และความแปรปรวน (variances) ที่ไม่คงที่ในแต่ละช่วงเวลาที่แตกต่างกัน

หาค่า D ด้วยวิธีการตรวจสอบอันดับความหยุดนิ่ง (Canova-Hansen,1995) ประมาณค่าของความแตกต่างตามฤดูกาล

ส่วนค่า p, q และ P,Q ใช้วิธีขยับตัวเลขไปทีละขั้นเรียกว่า (Stepwise Hyndman and Khandakar, 2018) หาแบบจำลองโอกาสที่เป็นไปได้ ดังตัวอย่าง

ARIMA(2,d,2) SARIMA(2,d,2)(1,D,1)

ARIMA(0,d,0) SARIMA(0,d,0)(0,D,0)

ARIMA(1,d,0) SARIMA(1,d,0)(1,D,0)

ARIMA(0,d,1) SARIMA(0,d,1)(0,D,1)

พิจารณาแบบจำลองด้วยค่า Akaike information criterion (AIC)

(Akaike, 1973) ได้เสนอเกณฑ์การคัดเลือกแบบจำลองขึ้นมาที่ให้ค่าพยากรณ์แม่นยำที่สุดเป็นเกณฑ์ที่สร้างจากการประมาณค่าความคาดเคลื่อนรวมเข้ากับข้อสนเทศ (Information) ของค่าสังเกตและใช้แนวคิดจากการหาค่าน้อยสุดของข้อสนเทศด้วยหลักการคูลแบล็ค-ไลเบอร์ (Kullback – Leibler, 1951)

$$AIC = \mathrm{n} \cdot \ln(\frac{\mathrm{SSE}}{\mathrm{n}}) + 2p$$

n คือ ขนาดตัวอย่าง

ln คือ ลอการิทึมฐาน e

SSE คือ ค่าความคลาดเคลื่อนกำลังสองของแบบจำลองการถดถอย

p คือ จำนวนพารามิเตอร์ของแบบจำลองการถดถอย

งานวิจัยนี้จะนำแบบจำลอง ARIMA(p,d,q) SARIMA(p,d,q)(P,D,Q) ที่ AIC มีค่าน้อยที่สุดไปใช้สร้างแบบจำลองต่อไป

2.5 เปรียบเทียบค่าความคาดเคลื่อนในการประเมินผลการพยากรณ์

เทียบความแตกต่างระหว่างข้อมูลจริงกับค่าพยากรณ์จากค่าความคลาดเคลื่อนที่เกิดขึ้น สำหรับงานวิจัยนี้จะเลือกนำตัวชี้วัดในการประเมินคือค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$$

n คือจำนวนตัวอย่าง

Y_t คือข้อมูลจริง ณ เวลา t

$\hat{Y}_t$ คือค่าพยากรณ์ของข้อมูล ณ เวลา t

2.6 งานวิจัยที่เกี่ยวข้อง

นิฉา แก้วหาวงษ์ (2547) ในงานวิจัยนี้ได้นำเสนอการพยากรณ์ข้อมูลอนุกรมเวลาโดยวิธีการทำให้เรียบแบบเอ็กซ์โปเนนเชียลและวิธีการของบ็อกซ์-เจนกินส์ของการพยากรณ์มูลค่าการส่งออกข้าว ยางพารา และมันสำปะหลังพบว่าแบบจำลอง ARIMA โดยวิธีการของบ็อกซ์-เจนกินส์เป็นแบบจำลองที่ดีที่สุดโดยให้ความคลาดเคลื่อนต่ำสุด

วรางคณา เรียนสุทธิ์ (2562) ในงานวิจัยนี้ได้นำเสนอเทคนิคพยากรณ์ที่เหมาะสมกับอนุกรมเวลาราคามะพร้าวด้วยวิธีบ็อกซ์-เจนกินส์ของการพยากรณ์ราคามะพร้าวพบว่าวิธีบ็อกซ์-เจนกินส์ที่มีแบบจำลอง AR(2) I(2) MA(2, 13, 15) ไม่มีพจน์ค่าคงตัวเป็นวิธีที่มีความถูกต้องและเหมาะสมมากที่สุด

หทัยชนก นานานอก (2553) ในงานวิจัยนี้ได้นำเสนอการพยากรณ์ยอดขายสินค้าเพื่อการวางแผนผลิตจากการเปรียบเทียบด้วยเทคนิคการพยากรณ์ 5 แบบ คือวิธีการพยากรณ์การหาค่าเฉลี่ยเคลื่อนที่ วิธีเทคนิควินเตอร์ วิธีทำให้เรียบแบบเอกซ์โปเนนเชียลชั้นเดียว วิธีทำให้เรียบแบบ

เอกซ์โปเนนเชียลสองชั้นและวิธีวิเคราะห์แนวโน้มเชิงเส้นพบว่าวิธีการพยากรณ์ที่ให้ค่าความคลาดเคลื่อนน้อยที่สุดคือวิธีวิเคราะห์แนวโน้มเชิงเส้น

Abdulwahed Salam, Abdelaaziz El Hibaoui (2018) ในงานวิจัยนี้ได้นำเสนอการเปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) สำหรับทำนายการใช้พลังงานไฟฟ้าเมืองเตโตอวนด้วยเทคนิคการพยากรณ์ 5 วิธี Linear regression, Decision tree, Random forest, Feedforward Neural network และ Supper vector พบว่าวิธี Random forest เป็นวิธีที่เหมาะสมที่สุด

Balpreet Singh, Pawan Kumar, Nonita Sharma and K P Sharma (2020) ในงานวิจัยนี้ได้นำเสนอการพยากรณ์ยอดขายในอนาคตของ Amazon.com, Inc. ด้วยเทคนิคการพยากรณ์ 4 แบบ คือ Holt-Winters Exponential Smoothing, Neural Network Autoregression Model, ARIMA (AutoRegressive Integrated Moving Average) และ SARIMA (Seasonal AutoRegressive Integrated Moving Average) พบว่า SARIMA ให้ผลลัพธ์ที่แม่นยำที่สุดเมื่อเปรียบเทียบกับวิธีอื่นๆ

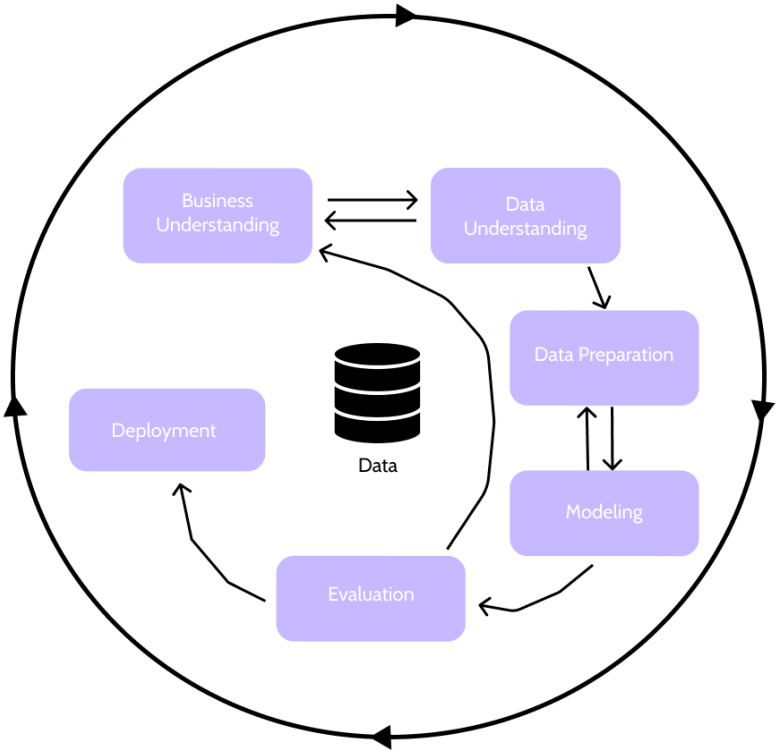
Takashi Tanizaki, Tomohiro Hoshino, Takeshi Shimmura and Takeshi Takenaka (2019) ในงานวิจัยนี้ได้นำเสนอการพยากรณ์จำนวนลูกค้าจะที่เข้ามาใช้บริการในร้านอาหารจากการเปรียบเทียบด้วยเทคนิคการพยากรณ์ 4 แบบ คือ Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression and Stepwise method พบว่าแต่ละวิธีการพยากรณ์ให้อัตราการพยากรณ์ (Forecast rate) จากการใช้เครื่องมือสร้างแบบจำลองการเรียนรู้และการวิเคราะห์ทางสถิติ ที่ชื่อว่า Azure Machine Learning และ SPSS ให้ค่าเฉลี่ยเกิน 85% ขึ้นไป

บทที่ 3
ระเบียบวิธีวิจัย

การศึกษาวิจัยครั้งนี้เป็นการวิจัยเชิงพยากรณ์ (Predictive) เป็นการวิจัยเพื่อที่จะนำผลของการพยากรณ์ข้อมูลไปใช้เป็นข้อมูลสำหรับการวางแผนบริหารจัดการทางด้านต่างๆ เพื่อหลีกเลี่ยงความเสี่ยงที่อาจเกิดขึ้นในอนาคตโดยใช้ข้อมูลในรูปแบบของอนุกรมเวลา

3.1 แนวทางการวิจัย

แนวทางการวิจัยนำหลักการ Cross-Industry Standard Process For Data Mining (CRISP-DM) ที่เป็นที่นิยมในการทำเหมืองข้อมูลมาประยุกต์ใช้



ภาพที่ 3.1 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

โดยนำข้อมูลอนุกรมเวลาที่แตกต่างกันมาทำการทดลองใช้กับระบบที่พัฒนาขึ้นมาเป็น 3 ประเภทของข้อมูลดังนี้

1. ข้อมูลยอดขายของร้านกาแฟ
2. ข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม
3. ข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3

3.1.1 การทำความเข้าใจกับธุรกิจ (Business Understanding)

3.1.1.1 กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ

กรณีนี้เป็นธุรกิจประเภทร้านค้า SME เราได้ศึกษาและนำโมเดลธุรกิจพื้นฐานคือ Business Model Canvas ที่จะช่วยให้เรามองเห็นรายละเอียดต่างๆของธุรกิจตัวเองได้อย่างดีไม่ว่าจะเป็นทั้งจุดเด่นจุดด้อยข้อได้เปรียบข้อเสียเปรียบกิจกรรมหลักของธุรกิจ หรือแม้กระทั่งคู่ค้าและลูกค้าของเราคือใครเพื่อที่จะมองเห็นจุดที่ต้องพัฒนาปรับปรุงให้ดีขึ้นดังภาพที่ 3.2

<table>
<tr>
<td rowspan="2">Key Partners
(คู่ค้าที่เกี่ยวข้อง)

- ผู้ขายวัตถุดิบ
- โรงเรียนและผู้ฝึกสอนบาริสต้า
- เจ้าของพื้นที่
- ผู้ขายเครื่องมือ อุปกรณ์</td>
<td>Key Activities
(กิจกรรมที่ต้องทำเพื่อให้โมเดลธุรกิจอยู่และไปต่อได้)
- การจัดหาวัตถุดิบ
- การพัฒนาเมนูใหม่ๆและพัฒนาสูตรเครื่องดื่มเดิมให้ดีขึ้น
- พัฒนาฝีมือบาริสต้า
- พัฒนาบริการดูแลลูกค้า</td>
<td rowspan="2">Value Propositions
(จุดขาย จุดเด่นที่เป็นคุณค่าของธุรกิจเรา)

- เมนูที่หลากหลาย
- คุณภาพของวัตถุดิบและเมล็ดกาแฟ
- ราคาเครื่องดื่มไม่แพงเมื่อเทียบกับคุณภาพ</td>
<td>Customer Relationships
(วิธีในการรักษาฐานลูกค้าเป็นแบบไหน)

- ให้ลูกค้าชิมกาแฟเมนูใหม่ๆ
- มีโปรโมชั่นสม่ำเสมอคืนกำไรให้ลูกค้า
- สร้างความสัมพันธ์ที่ดีกับลูกค้า สามารถจดจำลูกค้าประจำได้ว่าลูกค้าชอบแบบไหนต้องการอะไร</td>
<td rowspan="2">Customer Segments
(ลูกค้าเราคือใคร มีหน้าตาแบบไหน)

- คนที่มีรสนิยมชื่นชอบกาแฟ
- คนที่ชอบทานเครื่องดื่มและถ่ายรูป
- คนชอบทานเครื่องดื่มมีคุณภาพ</td>
</tr>
<tr>
<td>Key Resources
(ทรัพยากรที่จำเป็นที่ต้องมีในการดำเนิน

- สูตรกาแฟและเครื่องดื่มที่คิดขึ้นมาเอง
- การฝึกอบรมพนักงานที่มีทักษะ</td>
<td>Channels
(ช่องทางที่เข้าถึงลูกค้า)

- Social Media (Facebook, Instagram, Line)
- Call
- Food Delivery (Grab, Lineman, Robinhood)</td>
</tr>
<tr>
<td colspan="3">Cost Structure

- ค่าเช่าพื้นที่
- ค่าตกแต่งร้าน
- วัตถุดิบและอุปกรณ์
- ค่าเงินเดือนพนักงาน
- ค่าน้ำค่าไฟ
- ค่าประชาสัมพันธ์</td>
<td colspan="2">Revenue Streams (กระแสรายได้ของธุรกิจนี้)

- เครื่องดื่ม</td>
</tr>
</table>

ภาพที่ 3.2 Business Canvas ของร้านกาแฟที่ทำการวิจัย

นอกจากนั้นแล้วผู้วิจัยได้เข้าศึกษาเพิ่มเติมโดยการเข้าไปเป็นพนักงานในร้านเพื่อทำความเข้าใจและเพิ่มโอกาสเห็นปัญหาจริงที่สามารถกำหนดขอบเขตและนำข้อมูลมาวิเคราะห์เพื่อแก้ปัญหาได้พบว่าปัญหาหลักๆ ที่พบคือการประมาณการสั่งวัตถุดิบที่เก็บไว้ไม่ได้นาน เช่น นมสด

น้ำมะนาวสด หรือการผสมสูตรนมไม่เพียงพอต่อการขายต่อวันดังนั้นเราจึงนำขอบเขตของปัญหานี้มากำหนดขอบเขตในการวิเคราะห์ในงานวิจัยนี้

3.1.1.2 กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการ โครงการรถไฟฟ้ามหานครสายฉลองรัชธรรม

กรณีนี้เป็นองค์กรรัฐวิสาหกิจในสังกัดกระทรวงคมนาคมที่จัดขึ้นเพื่อดำเนินการระบบขนส่งมวลชนในรูปแบบรถไฟฟ้าในพื้นที่กรุงเทพมหานครและปริมณฑลมีเป้าหมายเพื่อบรรเทาความรุนแรงของปัญหาจราจรในประเทศเพื่ออำนวยความสะดวกให้ประชาชนเดินทางโดยไม่ต้องใช้รถยนต์ส่วนบุคคลดังนั้นเป้าหมายหลักของกรณีศึกษานี้ คือต้องการให้ประชาชนหันมาใช้ระบบขนส่งมวลชนให้มากขึ้นเพื่อลดปัญหาการจราจรลดค่าครองชีพให้กับประชาชน

3.1.1.3 กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3

กรณีนี้เป็นองค์กรของรัฐกระทรวงสาธารณสุขมีพันธกิจพัฒนาและอภิบาลระบบสุขภาพอย่างมีส่วนร่วมและยั่งยืน มีวิสัยทัศน์คือเป็นองค์กรหลักด้านสุขภาพที่รวมพลังสังคมเพื่อประชาชนสุขภาพดีจากสถานการณ์การแพร่ระบาดของโรคติดต่อเชื้อไวรัสโคโรนา 2019 ซึ่งมีการแพร่ระบาดทั่วโลกรวมถึงประเทศไทยส่งผลกระทบในวงกว้างด้านสาธารณสุข เศรษฐกิจ สังคม ความมั่นคงตลอดการดำรงชีพของประชาชนดังนั้นพันธกิจหลักขององค์กรนี้คือให้คำแนะนำวางแผนออกแนวทางการปฏิบัติ และมาตรการในการป้องกันความเสี่ยงจากโรคดังกล่าว

3.1.2 การทำความเข้าใจกับข้อมูล (Data Understanding)

3.1.2.1 กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ

รวบรวมข้อมูลที่เกี่ยวข้องกับการสร้างแบบจำลองพยากรณ์ยอดขายรายวันโดยเลือกใช้ข้อมูลยอดขายรายวันจากเครื่อง POS (Point of sale system) ของร้านค้าเพราะมีความน่าเชื่อถือและมีความเหมาะสมกับข้อมูล

วันที่	รหัสสินค้า	ชื่อสินค้า	กลุ่ม	หมวดสินค้า	ต้นทุนเฉลี่ย	ราคาขายเฉลี่ย	กำไรเฉลี่ย	จำนวนการขาย	ยอดก่อนลด	ต้นทุน	ส่วนลด	ราคาสุทธิ
03/08/2021		สตรอเบอร์รี่นมสดปั่น		โกยโชดา	0	45	45	1	45	0	0	45
03/08/2021		สตรอเบอร์รี่โยเกิร์ตปั่น		โกยโชดา	0	50	50	1	50	0	0	50
03/08/2021		บลูคูราโซ่		โกยโชดา	0	40	40	1	40	0	0	40
03/08/2021		น้ำผึ้งมะนาว		โกยโชดา	0	45	45	1	45	0	0	45
03/08/2021		มะนาวโชดา		โกยโชดา	0	35	35	1	35	0	0	35
02/08/2021		ช็อคโกแลต		โกยนม	0	45	45	2	90	0	0	90
02/08/2021		โกโก้ปั่นเย็น		โกยนม	0	40	40	1	40	0	0	40
02/08/2021		นมสดคาราเมล		โกยนม	0	35	35	1	35	0	0	35
02/08/2021		นมพาสเทล		โกยนม	0	40	40	1	40	0	0	40
02/08/2021		โอวัลตินไวบอลต์		โกยนม	0	30	30	1	30	0	0	30
02/08/2021		โอวัลติน		โกยนม	0	30	30	1	30	0	0	30
03/08/2021		โกโก้		โกยนม	0	38.75	38.75	8	310	0	0	310
03/08/2021		ช็อคโกแลต		โกยนม	0	41.25	41.25	4	165	0	0	165
03/08/2021		ช็อคโกแลตปั่นเย็น		โกยนม	0	48.33	48.33	3	145	0	0	145
03/08/2021		นมชมพู		โกยนม	0	35	35	1	35	0	0	35
03/08/2021		นมสดโอริโอ้ปั่น		โกยนม	0	40	40	3	120	0	0	120
02/08/2021		โกโก้		โกยนม	0	35	35	1	35	0	0	35
02/08/2021		ชามะนาว		โกยชา	0	42.5	42.5	4	170	0	0	170
02/08/2021		ชาแดสทีนมสด		โกยชา	0	35	35	1	35	0	0	35
02/08/2021		ชาไทยนมสด		โกยชา	0	40	40	16	640	0	0	640

ภาพที่ 3.3 ตัวอย่างข้อมูลยอดขายสินค้ารายวันที่จะนำมาวิเคราะห์

3.1.2.2 กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการ โครงการรถไฟฟ้ามหานครสายฉลองรัชธรรม

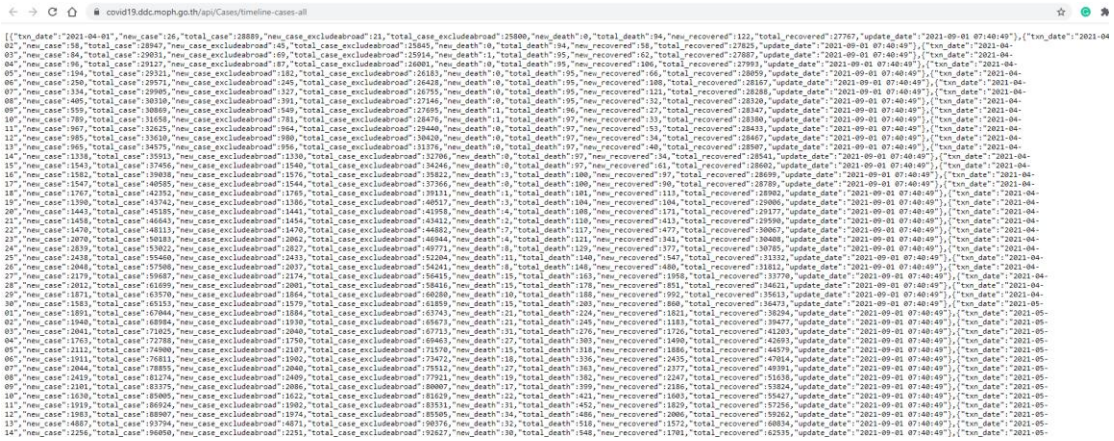
รวบรวมข้อมูลที่เกี่ยวข้องกับการสร้างแบบจำลองพยากรณ์จำนวนผู้โดยสารรวมรายเดือนโดยเลือกใช้ข้อมูลจากองค์กรการไฟฟ้าขนส่งมวลชนแห่งประเทศไทยเพราะมีความน่าเชื่อถือและมีการปรับปรุงข้อมูลให้อัปเดตอยู่เสมอ

เดือน	ปี	โครงการ	จำนวนผู้โดยสารรวม	จำนวนผู้โดยสารเฉลี่ยรายวัน	จำนวนผู้โดยสารเฉลี่ยรายวันธรรมดา	จำนวนผู้โดยสารเฉลี่ยรายวันหยุด
สิงหาคม	2559	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	524,472	20,172	20,492	19,568
กันยายน	2559	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	629,963	20,999	22,658	16,437
ตุลาคม	2559	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	650,772	20,993	24,212	15,140
พฤศจิกายน	2559	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	647,337	21,578	24,238	14,264
ธันวาคม	2559	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	593,981	19,161	23,048	12,093
มกราคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	634,552	20,469	24,282	13,538
กุมภาพันธ์	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	611,029	21,822	25,410	14,248
มีนาคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	719,225	23,201	25,447	16,741
เมษายน	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	631,343	21,045	25,678	15,750
พฤษภาคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	748,456	24,144	28,127	16,901
มิถุนายน	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	817,301	27,243	30,217	19,066
กรกฎาคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	823,504	26,565	31,501	18,749
สิงหาคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,331,334	42,946	47,655	31,436
กันยายน	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,407,258	46,909	52,322	34,278
ตุลาคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,376,904	45,897	53,861	32,140
พฤศจิกายน	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,395,555	46,519	52,042	31,330
ธันวาคม	2560	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,249,626	40,311	48,421	27,470

ภาพที่ 3.4 ตัวอย่างข้อมูลจำนวนผู้โดยสาร โครงการรถไฟฟ้ามหานครสายเฉลิมรัชธรรม ที่นำมาวิเคราะห์

3.1.2.3 กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3

รวบรวมข้อมูลที่เกี่ยวข้องกับการสร้างแบบจำลองพยากรณ์จำนวนผู้เสียชีวิตรายใหม่รายวันโดยเลือกใช้ข้อมูลจาก API ของกระทรวงสาธารณสุขเพราะมีความน่าเชื่อถือและมีการอัปเดตข้อมูลทุกวัน



ภาพที่ 3.5 ตัวอย่างข้อมูลจำนวนป่วย COVID-19 ระลอก 3 ที่จะนำมาวิเคราะห์

3.1.3 การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูลถือเป็นขั้นตอนที่ใช้เวลามากที่สุดในขั้นตอนทั้งหมดเพราะเป็นส่วนที่สำคัญที่ส่งผลต่อประสิทธิภาพในการพยากรณ์โดยการเตรียมข้อมูลประกอบด้วย

3.1.3.1 การคัดเลือกข้อมูล (Data Selection)

สำหรับงานวิจัยนี้เราจะเลือกข้อมูลที่เป็นตัวเลขมาใช้เนื่องจากเป็นการวิเคราะห์อนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่องตามรายละเอียดดังภาพที่ 3.6

	รูปแบบที่ 1: เทคนิค ARIMA, SARIMA		รูปแบบที่ 2: ใช้กับเทคนิค Multiple Linear Regression, Decision Tree Random Forest	
	Attirbute	**Target**	**Attirbute**	**Target**
ฟิลด์บังคับ	Date	Total	Date, ID	Total
กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ	เวลา (Date),	ยอดขาย (Total)	เวลา (Date), รหัสสินค้า (ID), อุณหภูมิ (Temperature), วันหยุด (Holiday), มีการจัดโปรโมชั่น (Promotion),	ยอดขาย (Total)
กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	เวลา (Date),	จำนวนผู้โดยสารรวม (Total)		
กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3			วันแถลง (Date), รหัสประเทศ (ID), จำนวนผู้ป่วยรายใหม่ (new_case), จำนวนผู้ป่วยรักษาหายรายใหม่ (new_recovered)	จำนวนผู้ป่วยตายรายใหม่ (Total)

ภาพที่ 3.6 การคัดเลือกข้อมูลสำหรับการสร้างแบบจำลองสำหรับกรณีศึกษาและเทคนิคต่าง ๆ

3.1.3.2 การกลั่นกรองข้อมูล (Data Cleaning)

ขั้นตอนนี้เนื่องจากข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม และข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3 ค่อนข้างสมบูรณ์แล้วดังนั้นเราจะทำขั้นตอนนี้เฉพาะข้อมูลยอดขายของร้านกาแฟที่ยังมีบางส่วนต้องกรองข้อมูลเพิ่มเติม

รูปแบบอนุกรมเวลาควรเป็นช่วงเวลาที่ต่อเนื่อง เช่น รายวัน รายเดือน รายปี ในกรณีศึกษานี้มีการพบข้อมูลที่หายไป (missing value) คือเป็นช่วงเวลาที่มีการปิดร้านค้าจึงได้มีการแทนที่ข้อมูลด้วยค่าเฉลี่ยของยอดขายภายในเดือนนั้นๆ มาทำการแทนที่และมีการกรองข้อมูลที่เป็นค่าผิดปกติก่อนที่จะนำไปสร้างแบบจำลองการพยากรณ์ คือรายการที่ไม่มียอดขายจริงเช่น ชื่อสินค้าคือ “รายการไม่รับช้อนส้อมพลาสติก”, “ใช้สิทธิโครงการรัฐ”, “ลด 50% สูงสุด 100.- เพียงใส่โค้ด "KBANK" ในหน้าชำระเงิน” หรือ “ของแถม” เป็นต้น

3.1.3.3 การแปลงรูปแบบข้อมูล (Data Transformation)

ขั้นตอนนี้สร้างข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์คือ

1. กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ

เพิ่มแอตทริบิวต์ที่อาจจะเพิ่มประสิทธิภาพในการวิเคราะห์เพิ่มเติม เช่น กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ วันหยุด (อ้างอิงวันหยุดตามประเพณีของสถาบันการเงิน), มีการจัดโปรโมชั่นหรือไม่, อุณหภูมิ (องศาเซลเซียส)

แปลงข้อมูลให้เป็น Numerical (ตัวเลข) เช่น วันหยุด (Holiday), การจัดโปรโมชั่น (Promotion) ให้อยู่ในรูปแบบของ {0 , 1} ดังตารางที่ 3.1 และแปลงข้อมูลวันที่ออกเป็นวันที่เท่าไหร่ของสัปดาห์ (วันจันทร์-อาทิตย์ 0-6), เดือน, ปี, วันที่ ดังตารางที่ 3.2 ซึ่งขั้นตอนดังกล่าวจะถูกปรับให้หลังจากนำข้อมูลเข้าระบบแล้ว

ตารางที่ 3.1 ตัวอย่างการแปลงแอตทริบิวต์วันหยุด (Holiday), การจัดโปรโมชั่น (Promotion)

IS Holiday	Promotion
0	0
0	0
0	1
0	0

ตารางที่ 3.2 ตัวอย่างการแปลงแอตทริบิวต์วันที่ (Date)

Date_dayofweek	Date_month	Date_year	Date_day
0	6	2020	22
3	3	2020	5

จัดกลุ่มข้อมูลสินค้าชินเดียวกันไว้เป็นข้อมูลชุดเดียวกัน เช่นข้อมูลสินค้าที่นำมาใช้บางรายการคือประเภทสินค้าชนิดเดียวกันแต่ชื่อสินค้าไม่ตรงกันดังตารางด้านล่าง

ตารางที่ 3.3 ตัวอย่างการจัดกลุ่มชื่อสินค้าก่อนนำเข้าวิเคราะห์

ID	ชื่อสินค้า
1	.อเมริกาโน่น้ำผึ้ง
1	*พรีเมียม อเมริกาโน่เย็น(AMERICANO)
1	*อเมริกาโน่
1	*อเมริกาโน่เย็น(AMERICANO)
1	*อเมริกาโน่เย็น*(AMERICANO)
1	พรีเมียม อเมริกาโน่เย็น(AMERICANO)
1	อเมริกาโน่

2. กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3 เพิ่มแอตทริบิวต์ ID เป็นรหัสประเทศเพื่อนำเข้าระบบ (ฟิลด์บังคับ) กรณีนี้จะใส่เป็นรหัส ID เดียวกัน และแปลงข้อมูลวันที่ตามตารางที่ 3.2

จากขั้นตอนในการเตรียมข้อมูลข้างต้น งานวิจัยนี้จะแบ่งรูปแบบไฟล์ข้อมูลเพื่อนำเข้าระบบเพื่อสร้างแบบจำลองและพยากรณ์ดังภาพที่ 3.7

	รูปแบบที่ 1: เทคนิค ARIMA, SARIMA	**รูปแบบที่ 2: ใช้กับเทคนิค Multiple Linear Regression, Decision Tree Random Forest**	
การแบ่งไฟล์ก่อนนำเข้าแบบจำลองและพยากรณ์	คอลัมภ์ข้อมูลใน ไฟล์ train	คอลัมภ์ข้อมูลใน ไฟล์ train	คอลัมภ์ข้อมูลใน ไฟล์ test
กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ	เวลา (Date), ยอดขาย (Total)	เวลา (Date), รหัสสินค้า (ID), อุณหภูมิ (Temperature), วันหยุด (Holiday), มีการจัดโปรโมชั่น (Promotion), ยอดขาย (Total)	เวลา (Date), รหัสสินค้า (ID), อุณหภูมิ (Temperature), วันหยุด (Holiday), มีการจัดโปรโมชั่น (Promotion),
กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานครสายฉลองรัชธรรม	เวลา (Date), จำนวนผู้โดยสารรวม (Total)		
กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3		วันแถลง (Date), รหัสประเทศ (ID), จำนวนผู้ป่วยรายใหม่ (new_case), จำนวนผู้ป่วยรักษาหายรายใหม่ (new_recovered), จำนวนผู้ป่วยตายรายใหม่ (Total)	วันแถลง (Date), รหัสประเทศ (ID), จำนวนผู้ป่วยรายใหม่ (new_case), จำนวนผู้ป่วยรักษาหายรายใหม่ (new_recovered)

ภาพที่ 3.7 คอลัมภ์ในข้อมูลไฟล์ train, test สำหรับการสร้างแบบจำลองสำหรับกรณีศึกษาและเทคนิคต่าง ๆ

และมีรูปแบบของข้อมูลก่อนนำเข้าแบบจำลองจริงหลังการแปลงข้อมูลผ่านระบบอัตโนมัติในแต่ละกรณีดังภาพที่ 3.8 - 3.11

Date	Total
3/1/2021	7,100.0000
3/2/2021	6,940.0000
3/3/2021	7,355.0000
3/4/2021	7,395.0000
3/5/2021	6,890.0000
3/6/2021	8,060.0000

ภาพที่ 3.8 รูปแบบของข้อมูลจริงก่อนนำเข้าสร้างแบบจำลองด้วยเทคนิค ARIMA, SARIMA กรณีศึกษาข้อมูลประเภทยอดขายของร้านกาแฟ

Date	Total
1/1/2562	1591763
2/1/2562	1443333
3/1/2562	1664735
4/1/2562	1458908
5/1/2562	1530913
6/1/2562	1675410

ภาพที่ 3.9 รูปแบบของข้อมูลจริงก่อนนำเข้าสร้างแบบจำลองด้วยเทคนิค ARIMA, SARIMA กรณีศึกษาข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม

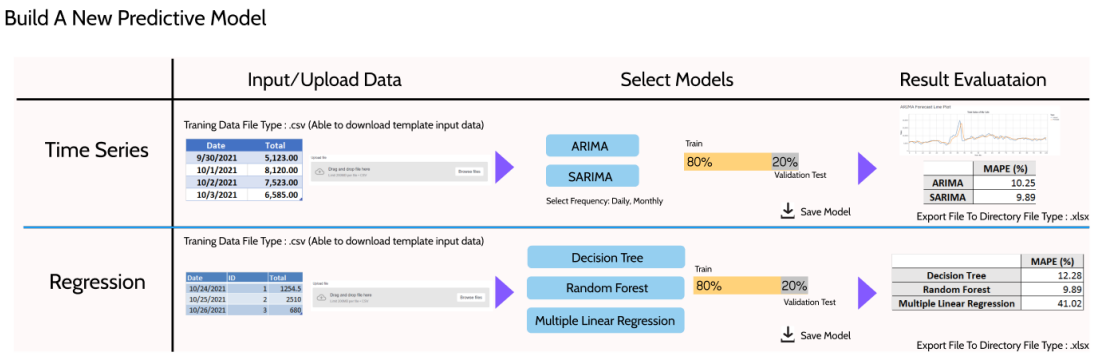
Temperature	IS Holiday	Promotion	ID	UnitPrice	Date_dayofweek	Date_month	Date_year	Date_day
29.1100	0	1	181	210	3	8	2021	5
30.3900	0	1	84	65	6	5	2021	30
28.0600	0	1	137	50	2	4	2021	28
32.1100	0	1	155	50	0	7	2021	5
31.0000	0	1	84	65	2	3	2021	24
29.1700	0	1	160	65	2	3	2021	3
32.2200	0	1	62	50	3	5	2021	6
28.9400	0	1	100	50	0	3	2021	22
31.0000	0	1	105	60	3	4	2021	22
32.1100	0	1	1	100	0	7	2021	5

ภาพที่ 3.10 รูปแบบของข้อมูลจริงก่อนนำเข้าสร้างแบบจำลองด้วยเทคนิค Multiple Linear Regression, Decision Tree, Random Forest กรณีศึกษาข้อมูลประเภทยอดขายของร้านกาแฟแบ่งตามรหัสสินค้า

ID	new_case	new_recovered	Date_dayofweek	Date_month	Date_year	Date_day
1	8685	3797	1	7	2021	13
1	11786	14738	1	9	2021	14
1	14956	17936	3	9	2021	2
1	6519	4148	2	7	2021	7
1	16031	15417	3	9	2021	9
1	2804	4143	6	6	2021	13
1	2331	4947	2	6	2021	16

ภาพที่ 3.11 รูปแบบของข้อมูลจริงก่อนนำเข้าสร้างแบบจำลองด้วยเทคนิค Multiple Linear Regression, Decision Tree, Random Forest กรณีศึกษาข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3

3.1.4 การสร้างแบบจำลอง (Modeling)



ภาพที่ 3.12 ขั้นตอนการทำงานของระบบเพื่อสร้างแบบจำลอง

ขั้นตอนนี้จะใช้ข้อมูลจากข้อ 3.1.3 มาทำการสร้างแบบจำลองโดยการแบ่งออกเป็น 2 กลุ่มหลักๆ คือ

3.1.4.1 อนุกรมเวลา (Time series)

ข้อมูลจะถูกนำเข้าเพื่อสร้างแบบจำลองด้วยวิธี ARIMA หรือ SARIMA โดยแบ่งข้อมูลจำนวน 80% เป็นข้อมูลสำหรับสร้างแบบจำลองและข้อมูลจำนวน 20% เป็นข้อมูลใช้เป็นข้อมูลทดสอบแบบจำลองที่สร้างขึ้นได้ใช้วิธีการ Split Test

การพิจารณารูปแบบของ ARIMA

กรณีที่เป็น Non Stationary series ที่มีแนวโน้มและมีความจำเป็นต้องทำ Regular Differencing หาค่า (d) อันดับความหยุดนิ่ง (จำนวนครั้งที่หาผลต่างเพื่อปรับอนุกรมเวลาให้เป็นอนุกรมเวลาที่ Stationary) ด้วยวิธี ADF test หา p, q ใช้วิธีขยับตัวเลขไปทีละขั้นหาแบบจำลองโอกาสที่เป็นไปได้และพิจารณาค่า AIC น้อยที่สุดเป็นแบบจำลองในการพยากรณ์

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=6710.951, Time=0.05 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=6675.320, Time=0.07 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=6662.405, Time=0.21 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=6708.995, Time=0.01 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=6657.294, Time=0.36 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=6653.920, Time=0.26 sec
 ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=6670.665, Time=0.06 sec
 ARIMA(3,1,1)(0,0,0)[0] intercept   : AIC=6654.851, Time=0.39 sec
 ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=inf, Time=0.61 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=6651.920, Time=0.49 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=6660.093, Time=0.18 sec
 ARIMA(1,1,3)(0,0,0)[0] intercept   : AIC=inf, Time=0.56 sec
 ARIMA(0,1,3)(0,0,0)[0] intercept   : AIC=6659.099, Time=0.34 sec
 ARIMA(2,1,3)(0,0,0)[0] intercept   : AIC=inf, Time=0.67 sec
 ARIMA(1,1,2)(0,0,0)[0]             : AIC=6651.986, Time=0.12 sec

Best model:  ARIMA(1,1,2)(0,0,0)[0] intercept
Total fit time: 4.385 seconds
```

ภาพที่ 3.13 ตัวอย่างการหารูปแบบ ARIMA (p,d,q) จากการพิจารณาค่า AIC

การพิจารณารูปแบบของ SARIMA

กรณีที่เป็น Non Stationary series ที่มีแนวโน้มและมีอิทธิพลของฤดูกาลมาเกี่ยวข้องมีความจำเป็นต้องทำ Seasonal Differencing หาค่า (D) อันดับความหยุดนิ่ง (จำนวนครั้งที่หาผลต่างเพื่อปรับอนุกรมเวลาให้เป็นอนุกรมเวลาที่ Stationary และไม่มีความเป็นฤดูกาล) ด้วยวิธี CH test หา p, q และ P,Q ใช้วิธีขยับตัวเลขไปทีละขั้นหาแบบจำลองโอกาสที่เป็นไปได้และแทนค่า s ด้วยคาบของฤดูกาล[5] รายวันแทนด้วย 7 รายเดือนแทนด้วย 12 งานวิจัยนี้จะมีข้อมูลที่นำมาทดลองเป็นลักษณะ รายวันและรายเดือน ซึ่งค่า s จะถูกปรับตามคาบของฤดูกาลที่ใช้และพิจารณาค่า AIC น้อยที่สุดเป็นแบบจำลองในการพยากรณ์

---

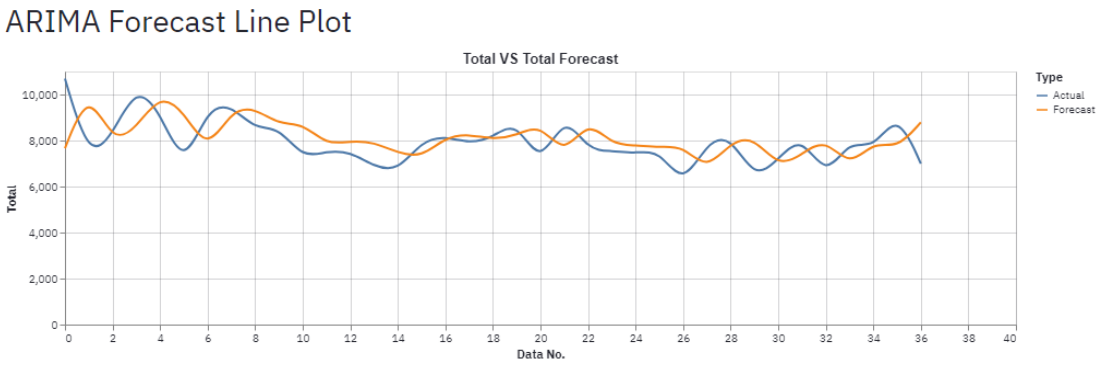
[5]pmdarima: ARIMA estimators for Python, © Copyright 2017-2021, Taylor G Smith จาก https://alkaline-ml.com/pmdarima/tips_and_tricks.html

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,1,1)[12]             : AIC=734.923, Time=0.04 sec
 ARIMA(0,1,0)(0,1,0)[12]             : AIC=747.992, Time=0.01 sec
 ARIMA(1,1,0)(1,1,0)[12]             : AIC=745.017, Time=0.03 sec
 ARIMA(0,1,1)(0,1,1)[12]             : AIC=738.959, Time=0.05 sec
 ARIMA(0,1,0)(1,1,1)[12]             : AIC=736.904, Time=0.05 sec
 ARIMA(0,1,0)(0,1,2)[12]             : AIC=736.450, Time=0.12 sec
 ARIMA(0,1,0)(1,1,0)[12]             : AIC=747.011, Time=0.03 sec
 ARIMA(0,1,0)(1,1,2)[12]             : AIC=737.523, Time=0.11 sec
 ARIMA(1,1,0)(0,1,1)[12]             : AIC=738.079, Time=0.05 sec
 ARIMA(1,1,1)(0,1,1)[12]             : AIC=740.918, Time=0.08 sec
 ARIMA(0,1,0)(0,1,1)[12] intercept   : AIC=737.099, Time=0.04 sec

Best model:  ARIMA(0,1,0)(0,1,1)[12]
Total fit time: 0.616 seconds
```

ภาพที่ 3.14 ตัวอย่างการหารูปแบบ ARIMA(p,d,q) SARIMA $(P,D,Q)_T$ หรือ ARIMA(p,d,q)x$(P,D,Q)_s$ จากการพิจารณาค่า AIC

หลังจากได้แบบจำลองที่เหมาะสมแล้วจะใช้ข้อมูล train ทั้งหมดเพื่อเข้าเรียนรู้สร้างแบบจำลองนำค่าพยากรณ์เทียบกับข้อมูล test ของแต่ละวิธี สามารถร่างกราฟเปรียบเทียบได้ดังตัวอย่างภาพที่ 3.15



ภาพที่ 3.15 กราฟเทียบข้อมูลพยากรณ์กับข้อมูลจริง (ข้อมูลทดสอบแบบจำลอง) ของเทคนิค ARIMA

3.1.4.2 สถิติข้อมูล/การวิเคราะห์เชิงถดถอย (Statistic/Regression Analysis)

ข้อมูลจะถูกนำเข้าเพื่อสร้างแบบจำลองด้วยวิธี Decision Tree, Random Forest และ Multiple Linear Regression โดยแบ่งข้อมูลจำนวน 80% เป็นข้อมูลสำหรับสร้างแบบจำลองและข้อมูลจำนวน 20% ใช้เป็นข้อมูลทดสอบโมเดลที่สร้างขึ้นได้ใช้วิธีการ Split Test แบบสุ่มข้อมูลก่อนแบ่ง

หลังจากได้แบบจำลองที่เหมาะสมแล้วจะใช้ข้อมูล train ทั้งหมดเพื่อเข้าเรียนรู้สร้างแบบจำลองและพยากรณ์ค่าออกมาดังตัวอย่างตารางที่ 3.4 และวัดค่าความคาดเคลื่อนจากข้อมูล test เทียบกับค่าที่พยากรณ์ได้ของแต่ละวิธีเพื่อเปรียบเทียบในลำดับถัดไป

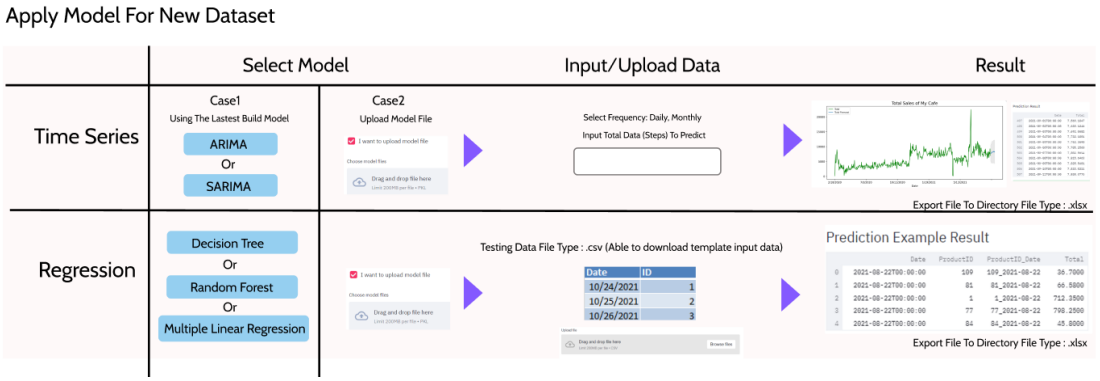
ตารางที่ 3.4 ตัวอย่างค่าพยากรณ์ที่ได้จากเทคนิค Random Forest (ข้อมูลทดสอบแบบจำลอง)

ID_Date	Actual	Forecast
1_2021-06-19	32	29.59
1_2021-07-16	67	94.02
1_2021-07-18	101	98.33
1_2021-08-29	264	274.96

3.1.5 การประเมินผล (Evaluation)

หลังจากที่แบ่งข้อมูลออกเป็น 2 ส่วนและใช้แบบจำลองที่สร้างได้จากข้อมูลจำนวน 80% มาทำการพยากรณ์ข้อมูลที่อยู่ในข้อมูลจำนวน 20% ที่เหลือนั้น และใช้ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE) แสดงค่าความคลาดเคลื่อนตามเปรียบเทียบด้วยเทคนิควิธีต่าง ๆ สามารถเลือกจัดเก็บแบบจำลองที่สนใจผ่านการ download ในรูปแบบของ .pkl ได้หรือ export ไฟล์ข้อมูลเปรียบเทียบดังตารางที่ 3.4 ในรูปแบบของ .xlsx

3.1.6 การปฏิบัติตามผลเสนอแนะ (Deployment)



ภาพที่ 3.16 ขั้นตอนการทำงานของระบบเพื่อพยากรณ์ข้อมูลใหม่

ขั้นตอนนี้หลังจากที่วิจัยและเปรียบเทียบค่าความคาดเคลื่อนของแต่ละเทคนิคแล้ว แบบจำลองใดเหมาะสมกับความต้องการพยากรณ์ข้อมูลในอนาคตสามารถเลือกใช้แบบจำลองได้ออกเป็น 2 รูปแบบคือ

3.1.6.1 เลือกใช้แบบจำลองที่ทำการสร้างล่าสุดในระบบ

รูปแบบการพยากรณ์นี้ระบบจะทำการจัดเก็บแบบจำลองล่าสุดอัตโนมัติ ผู้ใช้งานสามารถเลือกวิธีการพยากรณ์และทำการระบุคาบเวลาในอนาคตที่ต้องการพยากรณ์สำหรับเทคนิค ARIMA, SARIMA หรือ upload ข้อมูลสำหรับทดสอบสำหรับเทคนิค Decision Tree, Random Forest, Multiple Linear Regression ที่ต้องการพยากรณ์ข้อมูลในอนาคตระบบจะแสดงผลลัพธ์ออกมาในรูปแบบของกราฟและ/หรือตาราง สามารถ export ให้อยู่ในรูปแบบของ .xlsx ไฟล์ ให้สำหรับนำไปใช้งาน

3.1.6.2 เลือกใช้แบบจำลองที่ได้จัดเก็บไว้มาใช้พยากรณ์ข้อมูลใหม่

รูปแบบการพยากรณ์นี้ผู้ใช้งานสามารถ upload แบบจำลองที่เคยจัดเก็บไว้นำมาพยากรณ์ข้อมูลได้โดยทำการระบุคาบเวลาในอนาคตที่ต้องการพยากรณ์สำหรับเทคนิค ARIMA, SARIMA หรือ upload ข้อมูลสำหรับทดสอบสำหรับเทคนิค Decision Tree, Random Forest, Multiple Linear Regression ที่ต้องการพยากรณ์ข้อมูลในอนาคตระบบจะแสดงผลลัพธ์ออกมาในรูปแบบของกราฟและ/หรือตารางสามารถ export ให้อยู่ในรูปแบบของ .xlsx ไฟล์ ให้สำหรับนำไปใช้งาน

3.2 เครื่องมือที่ใช้ในการวิจัย

3.2.1 Python

เป็นภาษาที่ได้รับความนิยมมากที่สุดสำหรับการทำ Machine Learning, Deep Learning และ Data Science ในตอนนี้เพราะมีไวยากรณ์ที่เข้าใจง่ายไม่ซับซ้อน มี library มากมายให้ใช้ และมี Community ขนาดใหญ่สำหรับแลกเปลี่ยนความรู้และสอบถามได้เป็นอย่างดี

3.2.2 Library ที่ใช้สำหรับงานวิจัยนี้

Streamlit เป็น library ที่นำมาใช้พัฒนา Web Application เนื่องจากเป็นการใช้ภาษา Python ทำให้สามารถทำงานกับ Object อื่นๆได้เช่น Pandas, Matplotlib, NumPy เป็นต้นไม่จำเป็นต้องใช้ Library API อื่น (เช่น Flask, Django) เป็นตัวช่วยในการติดต่อ กับ HTML

Pandas เป็น library สำหรับจัดการข้อมูลต่างๆ เช่น การโหลดข้อมูลไฟล์ CSV เข้ามาแล้วแสดงข้อมูลให้ออกมาในรูปแบบ Row กับ Column คล้าย Table เราเรียกสิ่งนี้ว่า Data Frame ทำให้สามารถใช้งานข้อมูลเหล่านั้นได้ง่ายและนำไปใช้งานต่อได้อย่างหลากหลาย

NumPy เป็น library ที่ใช้ในการคำนวนทางคณิตศาสตร์ในภาษา Python ซึ่งภายในถูกเขียนด้วยภาษา C จึงทำงานได้เร็วและมีประสิทธิภาพโดย NumPy โดยทั่วไปจะเกี่ยวกับการจัดการข้อมูล Array ขนาดใหญ่และเมทริกซ์ขนาดใหญ่มและหลายมิติ

Matplotlib คือ library ที่ใช้สำหรับวาดกราฟที่ใช้งานได้หลากหลายยืดหยุ่น แสดงผลในรูปแบบ visualization เช่น กราฟเส้น, แผนภูมิแท่ง, แผนภูมิวงกลม เป็นต้น

Statsmodels คือ library ที่ช่วยให้สามารถใช้ Algorithm ที่สำคัญ ๆ อย่าง ANOVA และ ARIMA ที่ Machine Learning Library มาตรฐาน อย่าง Sci-kit Learn ไม่มีแล้วยังมีการแสดงรายละเอียด เช่น Features และ Metrics ของ Data ที่เป็นประโยชน์อีกมากมาย เช่น การใช้ Seasonal-Trend Decomposition

Scikit-learn คือ library สำหรับ Machine Learning ที่ได้รับความนิยมสุดๆ โดยมี algorithm ต่างๆ ทาง Machine Learning ให้ใช้งานง่ายและครบครัน เช่น Support Vector Machines (SVMs), Decision Tree, Random Forests, Gradient Boosting, Linear Regression เป็นต้น

Pmdarima เป็นโปรแกรมเสริม Python & Cython ของ Library ทางสถิติและการเรียนรู้ของเครื่อง (statsmodels และ scikit-learn) และดำเนินการการสร้างแบบจำลองกระบวนการ ARIMA ข้อดีของ Library นี้คือสามารถหาพารามิเตอร์ ARIMA (p, d, q) ที่ดีที่สุดสำหรับให้ได้เลย ในขณะที่ statsmodels จำเป็นจะต้องค้นหาพารามิเตอร์ที่เหมาะสมด้วยตนเอง

3.2.3 Anaconda

เป็น open-source ที่จะช่วยให้เราพัฒนาโปรแกรมทางด้าน data science และ machine learning ได้ง่ายยิ่งขึ้นมีการสร้าง Environment สำหรับการใช้ library ที่แตกต่างกันได้ง่าย รวมถึงมีการรวบรวมเครื่องมือต่างๆ ไว้ เช่น Jupyter Notebook, Code editor ที่ใช้สำหรับการเขียน python ที่นิยมก็จะมี VS Code, Spyder หรือ R Studio สำหรับการเขียนภาษา R เป็นต้น

บทที่ 4
ผลการศึกษา

การวิจัยครั้งนี้จะแบ่งผลการวัดประสิทธิภาพและความถูกต้องออกเป็น 2 ส่วนคือ การวัดผลประสิทธิภาพความถูกต้องของแบบจำลองและการวัดผลประสิทธิภาพความถูกต้องจากการนำแบบจำลองพยากรณ์ในอนาคตเทียบกับข้อมูลจริงที่เกิดขึ้นในอนาคต

4.1 ผลการวัดประสิทธิภาพความถูกต้องของแบบจำลอง

งานวิจัยนี้ใช้ข้อมูลสร้างแบบจำลองแบ่งตามกรณีศึกษา 3 กรณีคือ

1. กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ รายวันตั้งแต่เดือน มีนาคม ถึง สิงหาคม พ.ศ. 2564

2. กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานครสายฉลองรัชธรรม รายเดือนตั้งแต่เดือน สิงหาคม พ.ศ. 2559 ถึง ธันวาคม พ.ศ. 2563

3. กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3 รายวันตั้งแต่เดือน มิถุนายน ถึง กันยายน พ.ศ. 2564

นำข้อมูลแบบจำลองที่สร้างได้จากข้อมูลข้างต้นจำนวน 80% มาทำการพยากรณ์ ส่วนที่อยู่ในข้อมูลจำนวน 20% ที่เหลือนั้นใช้ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE) วัดผลค่าความคลาดเคลื่อนเพื่อเปรียบเทียบด้วยเทคนิคต่างๆ ได้ดังนี้

ตารางที่ 4.1 เปรียบเทียบค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของการพยากรณ์ด้วยกรณีศึกษาและเทคนิคในรูปแบบต่าง ๆ

	ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (Mean Absolute Percentage Error, MAPE)				
	รูปแบบที่ 1: เทคนิค ARIMA, SARIMA		**รูปแบบที่ 2: ใช้กับเทคนิค Multiple Linear Regression, Decision Tree Random Forest**		
เทคนิค	ARIMA	SARIMA	Decision Tree	Random Forest	Multiple Linear Regression
กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟ	14.22	14.22	57.54	54.51	156.86
กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานครสายฉลองรัชธรรม	35.15	25.90			
กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3			22.45	18.78	28.08

วิเคราะห์เปรียบเทียบค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของแบบจำลองทางอนุกรมเวลาพบว่าข้อมูลที่ใช้สร้างแบบจำลองด้วยเทคนิค ARIMA และ SARIMA ในครั้งนี้หากไม่มีแนวโน้มและอิทธิพลของฤดูกาลจะทำให้ผลพยากรณ์ที่ได้ มีค่าคาดเคลื่อนไม่ต่างกันซึ่งต่างจากกรณีข้อมูลที่มีอิทธิพลของฤดูกาลแบบจำลองด้วยเทคนิค SARIMA จะให้ผลค่าความคาดเคลื่อนน้อยกว่า

กรณีศึกษาประเภทข้อมูลยอดขายของร้านกาแฟมีจุดประสงค์เพื่อพยากรณ์ค่ายอดขายรวมรายวันจากค่าความคาดเคลื่อนดังตารางที่ 4.1 แสดงให้เห็นว่าสามารถเลือกใช้เพียงเทคนิค ARIMA ก็เพียงพอสำหรับการพยากรณ์ข้อมูลชุดใหม่

กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม มีจุดประสงค์เพื่อพยากรณ์จำนวนผู้โดยสารรวมรายเดือนจากค่าความคาดเคลื่อนดังตารางที่ 4.1 แสดงให้เห็นว่าการใช้เทคนิค SARIMA ในการสร้างแบบจำลอง มีค่าความคาดเคลื่อนน้อยกว่าซึ่งเหมาะสำหรับนำไปใช้กับการพยากรณ์ข้อมูลชุดใหม่

ส่วนวิเคราะห์เปรียบเทียบค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของแบบจำลองทางอนุกรมเวลาพบว่าข้อมูลที่ใช้สร้างแบบจำลองทางสถิติข้อมูล/การวิเคราะห์เชิงถดถอย (Decision Tree, Random Forest, Multiple Linear Regression) สำหรับกรณีศึกษาประเภทข้อมูลยอดขายของร้านกาแฟและกรณีศึกษาประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3 พบว่าจากค่าความคาดเคลื่อนมีลักษณะไปในทิศทางเดียวกันคือเทคนิค Random Forest ให้ค่าความคาดเคลื่อนน้อยที่สุดรองลงมาคือ Decision Tree และ Multiple Linear Regression ตามลำดับดังตารางที่ 4.1

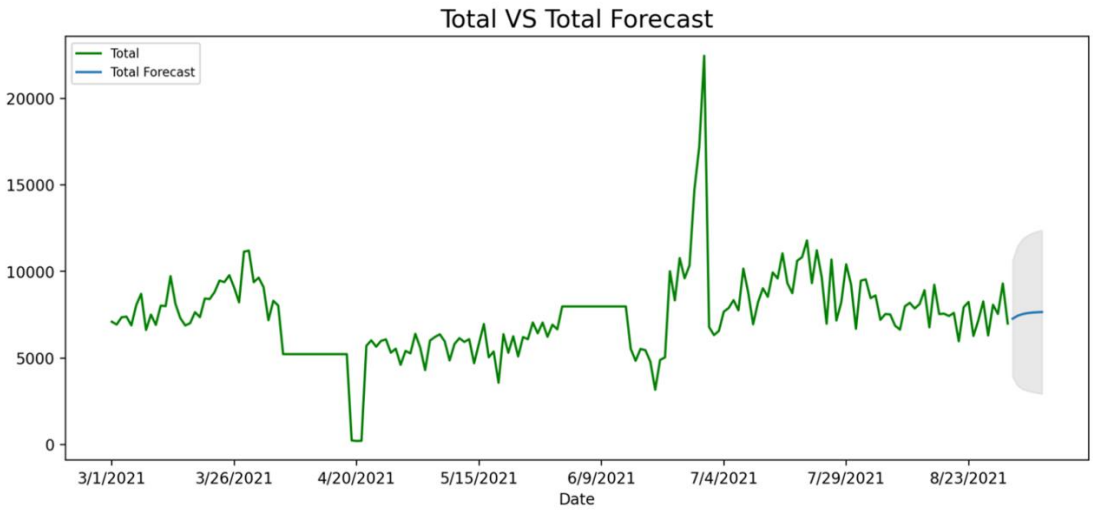
4.2 การวัดผลประสิทธิภาพความถูกต้องจากการนำแบบจำลองพยากรณ์ในอนาคตเทียบกับข้อมูลจริง

จากที่ได้ประทำการวัดผลประสิทธิภาพความถูกต้องของแบบจำลองแล้วจากข้อ 4.1 ขั้นตอนนี้จะเป็นการเลือกแบบจำลองที่เหมาะสมมาใช้พยากรณ์ข้อมูลอนาคต

4.2.1 กรณีศึกษาประเภทข้อมูลยอดขายของร้านกาแฟ

4.2.1.1 ตัวอย่างการนำแบบจำลองพยากรณ์ไปใช้พยากรณ์กับข้อมูลยอดขายรวมรายวัน

เลือกใช้เทคนิค ARIMA จากเทคนิค ARIMA, SARIMA สำหรับการสร้างแบบจำลองและพยากรณ์ข้อมูลอนาคต โดยการใส่จำนวนคาบเวลาของข้อมูลที่ต้องการพยากรณ์ในอนาคตระบบ เช่น ในกรณีนี้แทนค่าด้วยคาบเวลาเท่ากับ 7 วัน (วันที่ 1-7 กันยายน พ.ศ. 2564) โดยมีระบบแสดงผลการพยากรณ์ปรากฏดังภาพที่ 4.1 และ Export ข้อมูลพยากรณ์ให้ในรูปแบบของไฟล์ .xlsx ดังตัวอย่างตารางที่ 4.2



ภาพที่ 4.1 กราฟแสดงผลการพยากรณ์ข้อมูลยอดขายรวมรายวัน (1-7 กันยายน 2564) ด้วยเทคนิค ARIMA

ตารางที่ 4.2 ตัวอย่างข้อมูลพยากรณ์ยอดขายรวมรายวันด้วยเทคนิค ARIMA เมื่อกำหนดคาบเวลาเท่ากับ 7 สำหรับ Exportไฟล์ในรูปแบบ.xlsx

Future Step Forecast.	Total
1	7,276.209
2	7,441.324
3	7,538.273
4	7,595.198
5	7,628.622
6	7,648.247
7	7,659.77

เทียบกับข้อมูลจริงของยอดขายรวมรายวันของดังภาพที่ 4.2

วันที่	ยอดก่อนลด	ส่วนลด	ส่วนลดบิล	ยอดรวม	ค่าบริการ	ยอดก่อนภาษี	ภาษี	ลดท้ายบิล	ยอดปัดเศษ	รวมสุทธิ
01/09/2021	6,075	0	0	6,075	0	5,677.53	397.47	0	0	6,075
02/09/2021	6,385	0	0	6,385	0	5,967.28	417.72	0	0	6,385
03/09/2021	6,665	0	0	6,665	0	6,228.95	436.05	0	0	6,665
04/09/2021	5,530	0	0	5,530	0	5,168.24	361.76	114.75	2.75	5,418
05/09/2021	5,990	39	0	5,951	0	5,561.66	389.34	219.75	5.75	5,737
06/09/2021	6,060	0	0	6,060	0	5,663.57	396.43	274.5	6.5	5,792
07/09/2021	4,730	0	0	4,730	0	4,420.52	309.48	191.25	6.25	4,545
Total	41,435.00	39.00	0.00	41,396.00	0.00	38,687.75	2,708.25	800.25	21.25	40,617.00

ภาพที่ 4.2 ข้อมูลจริงของยอดขายรวมรายวัน วันที่ 1-7 กันยายน 2564

เมื่อหาค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของข้อมูลพยากรณ์เทียบกับข้อมูลจริง 7 วันได้ (อ้างอิงตารางที่ 4.2 และภาพที่ 4.2) เท่ากับ 39.23%

4.2.1.2 ตัวอย่างการนำแบบจำลองพยากรณ์ไปใช้พยากรณ์กับข้อมูลยอดขายรวมรายวันแบ่งตามรหัสสินค้า

เลือกใช้เทคนิค Random Forest จากเทคนิค (Decision Tree, Random Forest, Multiple Regression) สำหรับการสร้างแบบจำลองและพยากรณ์ข้อมูลอนาคตโดยการ upload ข้อมูลที่ต้องการพยากรณ์ดังตัวอย่างข้อมูลตามตารางที่ 4.3 ระบบแสดงข้อมูลจริงที่ทำการแปลงข้อมูลผ่านระบบดังภาพที่ 4.3 และแสดงผลการพยากรณ์และ Export ข้อมูลพยากรณ์ให้ในรูปแบบของไฟล์ .xlsx ปรากฏดังภาพที่ 4.4

ตารางที่ 4.3 ตัวอย่างข้อมูล Test สำหรับการพยากรณ์ข้อมูลใหม่ ด้วยเทคนิค Random Forest

Date	Temperature	IS Holiday	Promotion	ID	UnitPrice
9/1/2021	30	FALSE	TRUE	77	40
9/1/2021	30	FALSE	TRUE	141	50
9/1/2021	30	FALSE	TRUE	152	60
9/1/2021	30	FALSE	TRUE	158	55
9/1/2021	30	FALSE	TRUE	160	65
9/1/2021	30	FALSE	TRUE	111	45
9/1/2021	30	FALSE	TRUE	102	60

Temperature	IS Holiday	Promotion	ID	UnitPrice	Date_dayofweek	Date_month	Date_year	Date_day
30	0	1	77	40	2	9	2021	1
30	0	1	141	50	2	9	2021	1
30	0	1	152	60	2	9	2021	1
30	0	1	158	55	2	9	2021	1
30	0	1	160	65	2	9	2021	1
30	0	1	111	45	2	9	2021	1
30	0	1	102	60	2	9	2021	1

ภาพที่ 4.3 รูปแบบของข้อมูลจริงก่อนนำเข้าพยากรณ์ด้วย Random Forest กรณีศึกษาข้อมูลประเภทยอดขายของร้านกาแฟแบ่งตามรหัสสินค้า

Prediction Example Result

	Date	ID	ID_Date	Total
0	2021-09-01T00:00:00	77	77_2021-09-01	675.0000
1	2021-09-01T00:00:00	141	141_2021-09-01	506.8333
2	2021-09-01T00:00:00	152	152_2021-09-01	478.2500
3	2021-09-01T00:00:00	158	158_2021-09-01	491.1500
4	2021-09-01T00:00:00	160	160_2021-09-01	285.8000
5	2021-09-01T00:00:00	111	111_2021-09-01	435.5500
6	2021-09-01T00:00:00	102	102_2021-09-01	636.3000

ภาพที่ 4.4 ผลการพยากรณ์ข้อมูลยอดขายรวมรายรหัสสินค้าของวันที่ 1 กันยายน พ.ศ. 2564 ด้วยเทคนิค Random Forest สำหรับ Exportไฟล์ในรูปแบบ.xlsx

เทียบกับข้อมูลจริงของยอดขายรวมรายรหัสสินค้าของวันที่ 1 กันยายน พ.ศ. 2564 ดังภาพที่ 4.5

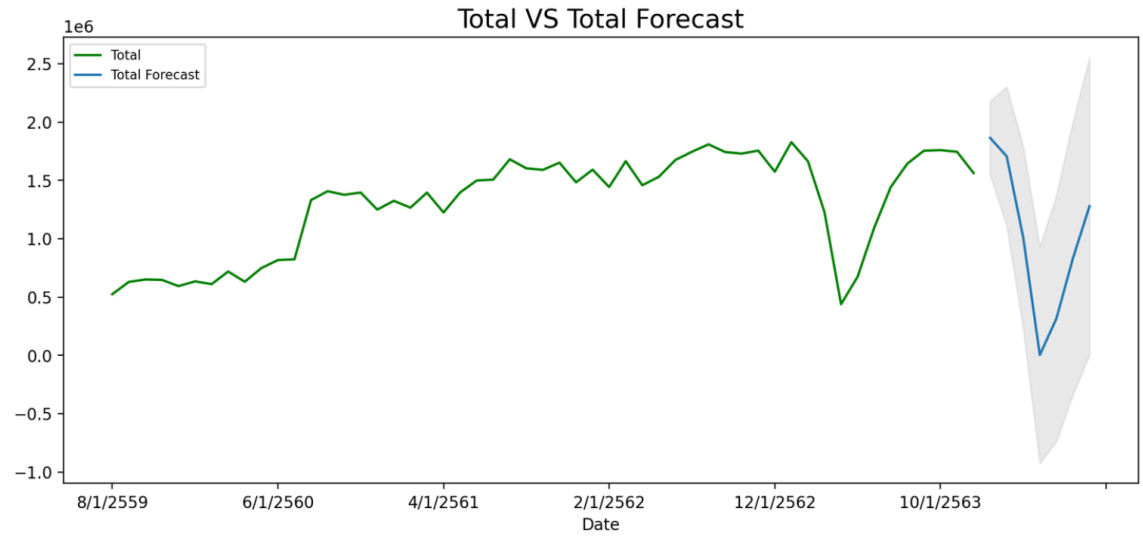
รหัสสินค้า	ชื่อสินค้า	กลุ่ม	หมวดสินค้า	ต้นทุนเฉลี่ย	ราคาขายเฉลี่ย	กำไรเฉลี่ย	จำนวนการขาย	ยอดก่อนลด	ต้นทุน	ส่วนลด	ราคาสุทธิ
77	เอส-เย็น		ไทย กาแฟ	0	41.43	41.43	14	580	0	0	580
141	พรีเมียม อเมริกาโน่เย็น		พรีเมียม เย็น	0	60	60	7	420	0	0	420
152	ลาเต้		ไทย กาแฟ	0	41	41	10	410	0	0	410
158	อเมริกาโน่		ไทย กาแฟ	0	40	40	10	400	0	0	400
160	อเมริกาโน่น้ำผึ้ง		ไทย กาแฟ	0	50	50	7	350	0	0	350
111	ชาไทยนมสด		ไทย ชา	0	36.67	36.67	9	330	0	0	330
102	คาปูชิโน่		ไทย กาแฟ	0	41.43	41.43	7	290	0	0	290

ภาพที่ 4.5 ข้อมูลจริงของยอดขายรวมรายรหัสสินค้าวันที่ 1 กันยายน 2564

เมื่อหาค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของข้อมูลพยากรณ์เทียบกับข้อมูลจริงของวันที่ 1 กันยายน 2564 จากตัวอย่างสินค้า 7 ตัวอย่างพบว่าได้ค่าความคาดเคลื่อน (อ้างอิงภาพที่ 4.4 และภาพที่ 4.5) เท่ากับ 35.18%

4.2.2 กรณีศึกษาประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม

เลือกใช้เทคนิค SARIMA จากเทคนิค ARIMA, SARIMA สำหรับการสร้างแบบจำลองและพยากรณ์ข้อมูลอนาคตโดยการใส่จำนวนคาบเวลาของข้อมูลที่ต้องการพยากรณ์ในอนาคตระบบ เช่น ในกรณีนี้แทนค่าด้วยคาบเวลาเท่ากับ 7 เดือน (เดือน มกราคม ถึง กรกฎาคม พ.ศ. 2564) โดยมีระบบแสดงผลการพยากรณ์ปรากฏดังภาพที่ 4.6 และ Export ข้อมูลพยากรณ์ให้ในรูปแบบของไฟล์ .xlsx ดังตัวอย่างตารางที่ 4.4



ภาพที่ 4.6 กราฟแสดงผลการพยากรณ์ข้อมูลจำนวนผู้ใช้บริการ (เดือน มกราคม ถึง เดือน กรกฎาคม พ.ศ. 2564) ด้วยเทคนิค SARIMA

ตารางที่ 4.4 ตัวอย่างข้อมูลพยากรณ์จำนวนผู้ใช้บริการรายเดือนด้วยเทคนิค SARIMA เมื่อกำหนดคาบเวลาเท่ากับ 7 สำหรับ Exportไฟล์ในรูปแบบ.xlsx

Future Step Forecast.	Total
1	1,864,581.63
2	1,706,402.79
3	1,008,853.91
4	3,758.38
5	315,255.02
6	833,198.08
7	1,279,183.94

กรณีนี้ผลลัพธ์ของจำนวนผู้ใช้บริการควรเป็นจำนวนเต็มก่อนนำมาใช้ให้ทำการปัดเลขเป็นจำนวนเต็มเพื่อนำไปเทียบกับข้อมูลจริงของจำนวนผู้ใช้บริการ ดังภาพที่ 4.7

เดือน	ปี	โครงการ	จำนวนผู้โดยสารรวม
มกราคม	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	874,003
กุมภาพันธ์	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,081,904
มีนาคม	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	1,519,467
เมษายน	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	793,625
พฤษภาคม	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	521,799
มิถุนายน	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	677,959
กรกฎาคม	2564	โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม	466,159

ภาพที่ 4.7 ข้อมูลจริงของจำนวนผู้ใช้บริการรายเดือน ตั้งแต่เดือน มกราคม ถึง กรกฎาคม 2564

เมื่อหาค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของข้อมูลพยากรณ์เทียบกับข้อมูลจริง 7 เดือนได้ (อ้างอิงตารางที่ 4.4 และภาพที่ 4.7) เท่ากับ 77.30%

4.2.3 กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3

เลือกใช้เทคนิค Random Forest จากเทคนิค (Decision Tree, Random Forest, Multiple Regression) สำหรับการสร้างแบบจำลองและพยากรณ์ข้อมูลอนาคตโดยการ upload ข้อมูลที่ต้องการพยากรณ์ดังตัวอย่างข้อมูลตามตารางที่ 4.5 ระบบจะทำการแปลงข้อมูลให้ดังภาพที่ 4.8 หลังจากนำข้อมูลเข้าระบบจะแสดงผลการพยากรณ์ประมวลผลด้วยแบบจำลองข้างต้นและสามารถ Export ข้อมูลพยากรณ์ให้ในรูปแบบของไฟล์ .xlsx ปรากฏดังภาพที่ 4.9

ตารางที่ 4.5 ตัวอย่างข้อมูล Test สำหรับการพยากรณ์ข้อมูลใหม่ ด้วยเทคนิค Random Forest

ID	Date	new_case	new_recovered
1	10/1/2021	11754	12473
1	10/2/2021	11375	13127
1	10/3/2021	10828	11894
1	10/4/2021	9930	12336
1	10/5/2021	9869	11152
1	10/6/2021	9866	10115

ID	new_case	new_recovered	Date_dayofweek	Date_month	Date_year	Date_day
1	11754	12473	4	10	2021	1
1	11375	13127	5	10	2021	2
1	10828	11894	6	10	2021	3
1	9930	12336	0	10	2021	4
1	9869	11152	1	10	2021	5
1	9866	10115	2	10	2021	6

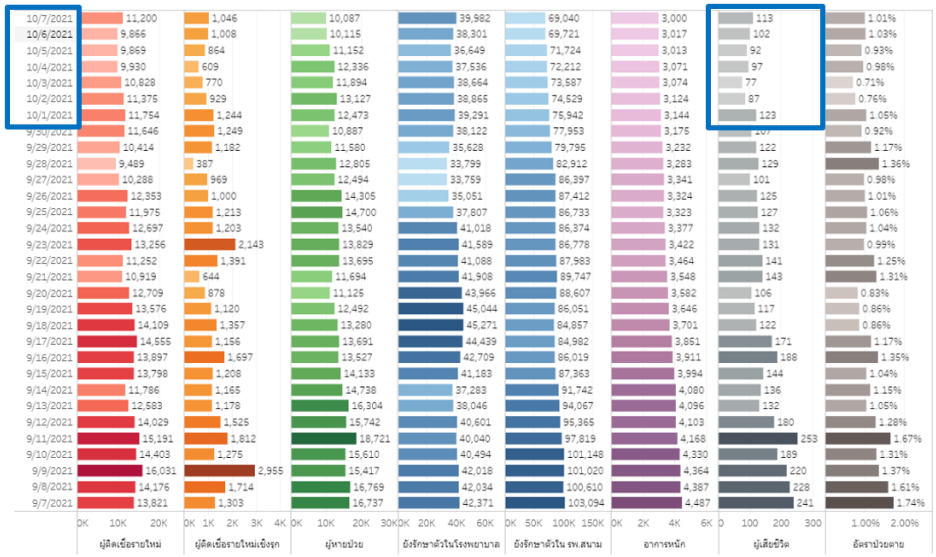
ภาพที่ 4.8 รูปแบบของข้อมูลจริงก่อนนำเข้าพยากรณ์ด้วย Random Forest กรณีศึกษาข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3

Prediction Example Result

	Date	ID	ID_Date	Total
0	2021-10-01T00:00:00	1	1_2021-10-01	132.0900
1	2021-10-02T00:00:00	1	1_2021-10-02	136.2400
2	2021-10-03T00:00:00	1	1_2021-10-03	131.8100
3	2021-10-04T00:00:00	1	1_2021-10-04	124.8400
4	2021-10-05T00:00:00	1	1_2021-10-05	127.0400
5	2021-10-06T00:00:00	1	1_2021-10-06	129.2400

ภาพที่ 4.9 ผลการพยากรณ์ข้อมูลผู้เสียชีวิตรายใหม่จากสถานการณ์ COVID-19 ระลอก 3 ด้วยเทคนิค Random Forest สำหรับ Exportไฟล์ในรูปแบบ.xlsx

เทียบกับข้อมูลจริงของเคส 30 วันย้อนหลังของรายงานสถานการณ์ COVID-19 ของวันที่ 1-6 ตุลาคม พ.ศ. 2564 ดังภาพที่ 4.10



ภาพที่ 4.10 ข้อมูลจริงของเคส 30 วันย้อนหลังของรายงานสถานการณ์ COVID-19

ที่มา: สรุปข้อมูลเคส 30 วันย้อนหลัง กรมควบคุมโรค กระทรวงสาธารณะสุข

เมื่อหาค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อนของข้อมูลพยากรณ์เทียบกับข้อมูลจริงของจำนวนผู้เสียชีวิตของรายงานสถานการณ์ COVID-19 วันที่ 1-6 ตุลาคม 2564 พบว่าได้ค่าความคาดเคลื่อน (อ้างอิงภาพที่ 4.9 และภาพที่ 4.10) เท่ากับ 37.69%

บทที่ 5
บทสรุปและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอเกี่ยวกับการพัฒนาระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง โดยผู้วิจัยได้วัดประสิทธิภาพของแต่ละตัวแบบจำลองแต่ละเทคนิค เพื่อนำผลลัพธ์มาเปรียบเทียบและสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

การพัฒนาระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่องสามารถประยุกต์ใช้การวิเคราะห์อนุกรมเวลาในทางปฏิบัติหลายอย่างเช่นกรณีศึกษาที่งานวิจัยนี้นำมาประยุกต์ใช้ คือ

1. กรณีศึกษาเป็นประเภทข้อมูลยอดขายของร้านกาแฟรายวันตั้งแต่เดือน มีนาคม ถึง สิงหาคม พ.ศ. 2564

2. กรณีศึกษาเป็นประเภทข้อมูลจำนวนผู้ใช้บริการโครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรมรายเดือนตั้งแต่เดือน สิงหาคม พ.ศ. 2559 ถึง ธันวาคม พ.ศ. 2563

3. กรณีศึกษาเป็นประเภทข้อมูลรายงานสถานการณ์ COVID-19 ระลอก 3 รายวันตั้งแต่เดือน มิถุนายน ถึง กันยายน พ.ศ. 2564

โดยใช้เทคนิคที่หลากหลายได้แก่ เทคนิค ARIMA, SARIMA, Decision Tree, Random Forest และ Multiple Linear Regression

จากผลการทดลองพิจารณาประสิทธิภาพของแบบจำลองเปรียบเทียบเทคนิค ARIMA กับ SARIMA เมื่อนำข้อมูลกรณีศึกษาของยอดขายรวม (รายวัน) ของร้านกาแฟและจำนวนผู้ใช้บริการ (รายเดือน) โครงการรถไฟฟ้ามหานคร สายฉลองรัชธรรม พบว่าหากข้อมูลมีการเคลื่อนไหวแนวโน้มที่ไม่เสถียรและไม่มีอิทธิพลตามฤดูกาลผลของการพยากรณ์จะให้ค่า MAPE ใกล้เคียงกันหรือแทบไม่ต่างกันเลย เนื่องจากพารามิเตอร์ที่ทำผลของเกณฑ์ในการคัดเลือกแบบจำลอง (AIC ต่ำที่สุด) สำหรับเทคนิคทั้งสองข้างต้นมีค่าเดียวกัน ในทางตรงกันข้ามหากข้อมูลมีแนวโน้มอิทธิพลตามฤดูกาลผลของการพยากรณ์ของเทคนิค SARIMA จะให้ผลดีกว่าโดยให้ค่า MAPE ต่ำกว่าเทคนิค ARIMA และเมื่อพิจารณาประสิทธิภาพของแบบจำลองเปรียบเทียบเทคนิค

Decision Tree, Random Forest และ Multiple Linear Regression เมื่อนำข้อมูลกรณีศึกษาของยอดขายรวมของแต่ละรายสินค้า (รายวัน) ของร้านกาแฟและจำนวนผู้ป่วยและผู้เสียชีวิตจากสถานการณ์ COVID-19 ระลอก 3 (รายวัน) พบว่าเทคนิค Random Forest พยากรณ์ยอดขายตามรายการสินค้าและจำนวนผู้เสียชีวิตรายใหม่ของสถานการณ์ COVID-19 ให้ผลค่าความเคลื่อนไปในทิศทางเดียวกันโดยให้ค่า MAPE ต่ำที่สุดรองลงมาคือเทคนิค Decision Tree และ Multiple Linear Regression ตามลำดับ สอดคล้องกับผลการศึกษาของ Abdulwahed Salam and Abdelaaziz El Hibaoui (2018) ซึ่งศึกษาเปรียบเทียบวิธีพยากรณ์ 5 วิธี Linear regression, Decision tree, Random forest, Feedforward Neural network และ Supper vector ของการพยากรณ์ปริมาณการใช้ไฟฟ้าของเมืองเตโตอวนประเทศโมร็อกโกพบว่าวิธี Random Forest เป็นวิธีที่เหมาะสมที่สุด

5.2 อภิปรายผลการวิจัย

จากการตั้งสมมติฐานที่ว่าค่าการพยากรณ์ควรมีความใกล้เคียงและมีค่า MAPE ไม่น้อยกว่า 50% แต่เมื่อนำข้อมูลสร้างแบบจำลองและวัดประสิทธิภาพจากค่าความคาดเคลื่อนพบว่าค่า MAPE ที่ได้ส่วนใหญ่ให้ผลตามสมมติฐานที่ตั้งไว้ แต่ก็มีบางกรณีที่ค่า MAPE สูงกว่าสมมติฐาน ผู้วิจัยได้ศึกษาเพิ่มเติมว่าอะไรเป็นอาจเป็นเหตุที่ทำให้ประสิทธิภาพของการพยากรณ์ไม่ดีเท่าที่ควร หลังจากได้ไปศึกษารายละเอียดข้อมูลและสภาพแวดล้อมหรือปัจจัยที่ส่งผลกระทบของข้อมูล ได้พบข้อสังเกตที่น่าสนใจ คือช่วงเวลาของข้อมูลที่นำมาวิจัยและเทียบผลการพยากรณ์นั้นมีความทิศทางที่ไม่เสถียรจากสถานการณ์ COVID-19 เช่น การขอความร่วมมือประชาชนงดออกจากเคหสถาน หรือ การออกแนวทางปฏิบัติด้านสาธารณสุขเพื่อป้องกันการแพร่ระบาดของเชื้อ COVID-19 สำหรับห้างสรรพสินค้า ศูนย์การค้า คอมมูนิตี้มอลล์ ร้านอาหาร แหล่งท่องเที่ยว หรือสถานที่ท่องเที่ยว เป็นต้น หรือแม้กระทั่งปัจจัยทางด้านสิ่งแวดล้อม สภาพภูมิอากาศ ตัวอย่างเช่น มีเหตุการณ์พายุเข้าไทยทำให้ฝนตกหนักช่วงเวลาสั้นๆ ส่งผลให้คนงดออกเดินทางหรือออกมาจับจ่ายใช้สอยซึ่งส่งผลกระทบโดยตรงกับข้อมูลที่นำมาศึกษาวิจัย

5.3 ข้อเสนอแนะ

5.3.1 ข้อมูลที่นำมาศึกษาวิจัยควรมีความสมบูรณ์ถูกต้องและมีความน่าเชื่อถือ เช่น การเตรียมสร้างรูปแบบการบันทึกของข้อมูล เช่น การกำหนดรหัสสินค้าให้ตรงกับสินค้าและราคาขายเพื่อที่ได้ข้อมูลที่สมบูรณ์ลดการเตรียมข้อมูลก่อนการนำเข้าสร้างแบบจำลอง

5.3.2 ควรแบ่งข้อมูลของคาบเวลามาสร้างแบบจำลองให้เหมาะสมตามจุดประสงค์ของการพยากรณ์ เช่นใช้ปัจจัยพิจารณาร่วมคือหากมีสถานการณ์ล็อกดาวน์ หรือการขอความร่วมมือประชาชนงดออกจากเคหสถาน ที่ทำให้ยอดขาย หรือจำนวนผู้ใช้บริการในช่วงเวลานั้นๆ ไม่ปกติ หากเรามีจุดประสงค์ว่าอยากทราบว่าถ้าช่วงเวลาปกติจะมียอดขายหรือจำนวนผู้ใช้บริการประมาณเท่าไหร่ ลองปรับเลือกข้อมูลยอดขายหรือจำนวนผู้ใช้บริการช่วงหลังจากคลายล็อกดาวน์ มาทำการพยากรณ์ หรืออาจจะใช้การวิเคราะห์เพิ่มเติมว่าคาบเวลาใดที่เหมาะสมกับการสร้างแบบจำลองเช่น 1 หรือ 3 เดือนก็อาจเพียงพอกับการพยากรณ์ข้อมูลรายวัน เป็นต้น

5.3.3 ควรมีการพิจารณาประเมินผลที่ได้จากการพิจารณาอย่าสม่ำเสมอ ปรับเพิ่มลดตัวแปรอิสระเพิ่มเติมที่อาจช่วยเพิ่มประสิทธิภาพของการสร้างแบบจำลองมากขึ้น หรือลองนำเทคนิคใหม่ๆ ที่ถูกพัฒนามาปรับใช้ให้มีประสิทธิภาพมากขึ้น

5.3.4 นอกจากตัวระบบที่ช่วยพยากรณ์เบื้องต้นแล้ว ควรมีการพิจารณาปัจจัยอื่นๆที่อาจส่งผลกับค่าพยากรณ์เพิ่มเติมร่วมด้วยเช่น ข้อมูลทางเศรษฐกิจ, แผนการพัฒนาผลิตภัณฑ์และรูปแบบทางการตลาดร่วมด้วย

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

การพยากรณ์.(2564). สืบค้น 27 กันยายน 2564, จาก http://119.46.166.126/digitalschool/p5/ma5_1/lesson5/content1/more/page2.php

ดร.สุทิน ชนะบุญ. (2560). สถิติการวิเคราะห์ข้อมูลในงานวิจัยเบื้องต้น บทที่ 6 การวิเคราะห์เชิงอนุมาน. R2R สำนักงานสาธารณสุขจังหวัดขอนแก่น ปี 2560. น.148-160

ธนัท จรณะสมบูรณ์. (2561). การทำนายการซื้อซ้ำของผู้ซื้อโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร. ปริญญานิพนธ์ วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีนครินทรวิโรฒ.

นิฉา แก้วหาวงษ์. (2547). การศึกษาเปรียบเทียบการพยากรณ์ข้อมูลอนุกรมเวลาโดยวิธีการทำให้เรียบแบบเอ็กซ์โปเนเนเชียลและวิธีการของบ็อกซ์-เจนกินส์: กรณีศึกษาการพยากรณ์มูลค่าการส่งออกข้าว ยางพารา และมันสำปะหลัง. วารสารวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 1 ฉบับที่ 2. (น.73-90).

ปณิธี ธัมมวิจยะ. (2560) การพยากรณ์โรคโดยการวิเคราะห์ข้อมูลอนุกรมเวลา. อมรมพัฒนาศักยภาพบุคลากรด้านการพยากรณ์โรค กองยุทธศาสตร์และแผนงาน สำนักงานปลัดกระทรวงสาธารณสุข. สืบค้น 30 กันยายน 2564, จาก https://bps.moph.go.th/new_bps/sites/default/files/ime%20Series%20Analysis_2017.pdf

ปริมาณผู้โดยสารรถไฟฟ้ามหานคร. (2564). สืบค้น 5 ตุลาคม 2564, จาก https://data.go.th/dataset/mrta-crmk

ปิยมาส กล้าแข็ง. (2561). การประยุกต์ใช้เทคนิคการพยากรณ์ เพื่อการจัดการสินค้าคงคลัง. วิทยานิพนธ์ ปริญญาวิทยาศาสตร์มหาบัณฑิต สาขาวิชาการจัดการซัพพลายเชนธุรกิจ มหาวิทยาลัยราชภัฏสวนสุนันทา.

ผศ.ดร. เฉลิมพล. (2562). การพยากรณ์ทางอนุกรมเวลา (Time Series Forecasting). สืบค้น 30 กันยายน 2564, จาก https://cj007blog.files.wordpress.com/2020/04/07-time-series-forecasting.pdf

วรางคณา เรียนสุทธิ์. (2562). การพยากรณ์ราคามะพร้าวด้วยวิธีบ็อกซ์-เจนกินส์.วารสารวิชาการ มทร.สุวรรณภูมิ (RMUTSB Acad. J.) ปีที่ 7. (น. 87 – 100). มหาวิทยาลัยทักษิณ.

วันหยุดตามประเพณีของสถาบันการเงิน ประจำปี. (2564). สืบค้น 14 มิถุนายน 2564, จาก https://www.bot.or.th/Thai/FinancialInstitutions/FIholiday/Pages/2021.aspx

สถานการณ์ผู้ติดเชื้อ COVID-19 อัพเดทรายวัน รูปแบบ API(Json/CSV Data Format). (2564). สืบค้น 28 กันยายน 2564, จาก https://covid19.ddc.moph.go.th/

หทัยชนก นานานอก. (2553). ศึกษาการพยากรณ์ยอดขายเพื่อวางแผนการผลิตของสินค้าในอุตสาหกรรมขนาดเล็กแห่งหนึ่ง. ปัญหาพิเศษอุตสาหกรรมศาสตร์มหาบัณฑิต. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

Fanny Indradjaja. (2562). Quantitative Analysis. สืบค้น 27 กันยายน 2564, จาก https://slideplayer.in.th/slide/16724291/

ภาษาต่างประเทศ

Akaike, H. (1973). Information Theory as an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., Second International Symposium on Information Theory, Akademiai Kiado, Budapest, pp.267-281.

A Salam, A El Hibaoui. (2018). Comparison of Machine Learning Algorithms for the Power Consumption Prediction : - Case Study of Tetouan city-. 2018 6th International Renewable and Sustainable Energy Conference (IRSEC). pp 1210

Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8

B. Singh P. Kumar N. Sharma and K P Sharma. (2020). Sales Forecast for Amazon Sales with Time Series Modeling. 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T). pp 38-43.

Canova, F. and Hansen, Bruce E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. Journal of Business & Economic Statistics, 13(3), pp. 237-252.

Georgios D. (2019). Decision Tree Regressor explained in depth. Retrieved September 19, 2021, from https://gdcoder.com/decision-tree-regressor-explained-in-depth/

Kullback, S., &Leibler, R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics,22, 79–86.

Taylor G Smith. (2017). pmdarima: ARIMA estimators for Python. Retrieved August 3,2021, from https://alkaline-ml.com/pmdarima/tips_and_tricks.html

T. Tanizaki, T. Hoshino, T. Shimmura and T. Takenaka. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. Procedia CIRP. vol. 79. pp. 679-683.

Weather Underground. (2020). Bangkok, Thailand Weather History. Retrieved June 14,2021 from https://www.wunderground.com/history/monthly/th/bangkok

ภาคผนวก

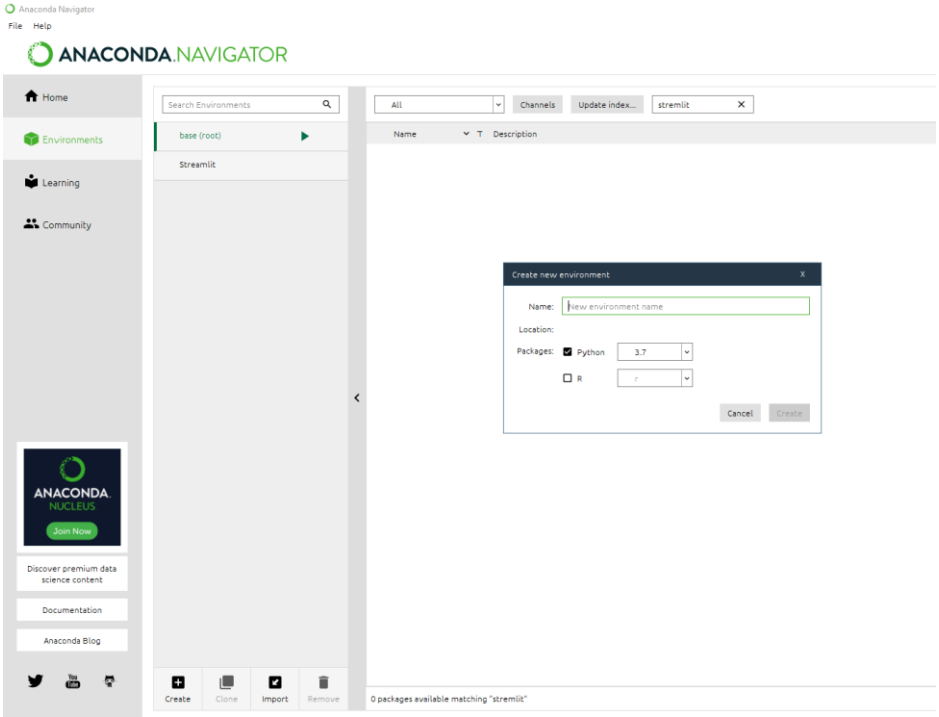
ภาคผนวก ก

ขั้นตอนการพัฒนาระบบ

ขั้นตอนการพัฒนาระบบ

1. ดาวน์โหลด Anaconda https://www.anaconda.com/products/individual และติดตั้ง ให้เลือก bit ตาม Windows ที่เราลง

2. กด Create Environment เลือก package type “Python”

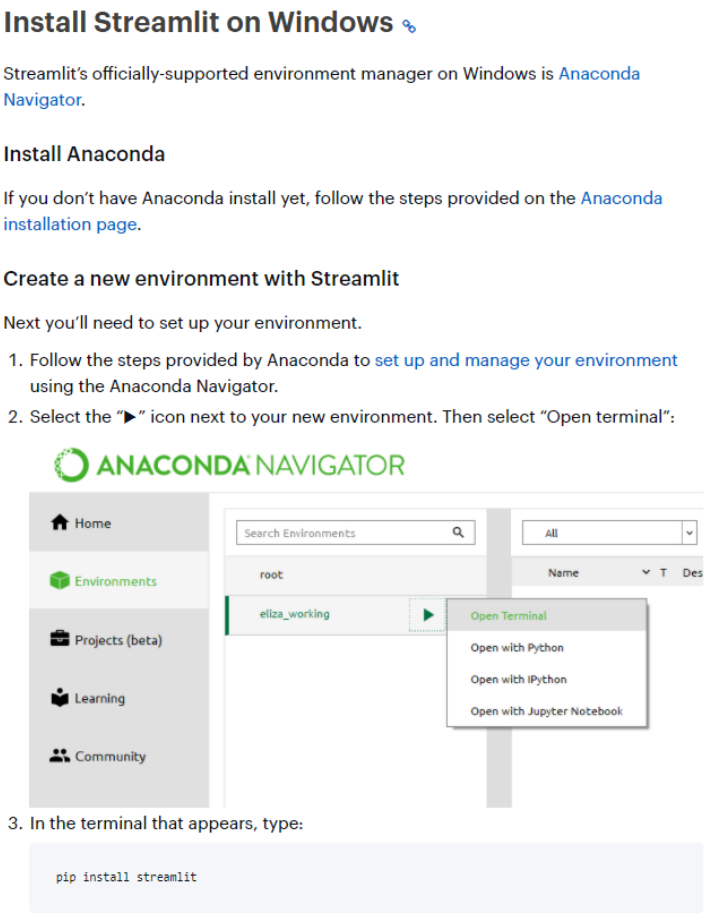


3. เปิด environment ที่สร้างไว้ แล้วทำการลง library ต่าง ๆ ที่เราใช้งาน

ตัวอย่าง การ Install Streamlit

Streamlit เป็น Python library ที่ถูกนำมาใช้พัฒนา Web Application สามารถทำงานกับ Library อื่นๆที่เป็นภาษา Python ด้วยกันได้ เช่น Pandas, Numpy, Matplotlib หรือ Scikit-learn เป็นต้น

https://docs.streamlit.io/en/stable/troubleshooting/clean-install.html#install-streamlit-on-windows หรือสำหรับ OS อื่นๆ https://docs.continuum.io/anaconda/install/



Install Streamlit on Windows

Streamlit's officially-supported environment manager on Windows is Anaconda Navigator.

Install Anaconda

If you don't have Anaconda install yet, follow the steps provided on the Anaconda installation page.

Create a new environment with Streamlit

Next you'll need to set up your environment.

1. Follow the steps provided by Anaconda to set up and manage your environment using the Anaconda Navigator.
2. Select the "▶" icon next to your new environment. Then select "Open terminal":

3. In the terminal that appears, type:

```
pip install streamlit
```

สำหรับ Library อื่น ๆ สามารถดูขั้นตอนการ install ได้จาก documentation website ของ Library นั้นๆ เช่น Scikit learn, pmdarima, pandas, numpy, altair, matplotlib, statsmodels เป็นต้น

4. พัฒนาระบบด้วยการเขียน code python ตัวอย่าง code

4.1 import library

```python
import streamlit as st
import numpy as np
import pandas as pd
import seaborn as sns
import altair as alt
import matplotlib
import matplotlib.pyplot as plt
#from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn import model_selection
#from sklearn.preprocessing import LabelEncoder
# evaluate an ARIMA model using a walk-forward validation
from pandas import read_csv
from datetime import datetime
from matplotlib import pyplot
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.arima_model import ARIMAResults
from math import sqrt
from pandas import DataFrame
# Compute Seasonal Index
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.metrics import median_absolute_error, mean_squared_log_error
from sklearn.metrics import mean_absolute_percentage_error
import csv
#%matplotlib inline

import pandas_profiling

from datetime import datetime
import statsmodels.api as sm
from sklearn import metrics
#from sklearn.cross_validation import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.svm import SVC, LinearSVC
```

4.2 เขียน code ด้วยภาษา Python ตัวอย่าง code ด้านล่างคือการสร้าง model ของเทคนิค ARIMA

```
def parser(x):⊟
def arima(data):
    if data is not None:
        #st.success("Data successfully loaded")
        data_check = pd.read_csv(data.name)
        data_check.to_csv('train_timeseries.csv', index = False)
        #st.write(data_check.columns[0],data_check.columns[1])

        if 'Date' == data_check.columns[0] and 'Total' == data_check.columns[1]:

            d = pd.read_csv(data.name)

            # Create Training and Test
            size = int(len(d.Total) * 0.80)
            #train, test = X[0:size], X[size:len(d.Total)]
            train = d.Total[0:size]
            test = d.Total[size:len(d.Total)]
            test_date = d.Date[size:len(d.Date)].reset_index()


            testt = test.reset_index()
            kpss_diffs = ndiffs(train, alpha=0.05, test='kpss', max_d=6)
            adf_diffs = ndiffs(train, alpha=0.05, test='adf', max_d=6)
            n_diffs = max(adf_diffs, kpss_diffs)


            model = pm.auto_arima(train, start_p=0, start_q=0,
                      test='adf',       # use adftest to find optimal 'd'
                      max_p=3, max_q=3, # maximum p and q
                      d=n_diffs,           # let model determine 'd'
                      seasonal=False,   # No Seasonality
                      start_P=0,
                      trace=True,
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=True)

            tup = model.order

            # *******  save the model to disk
            filename = "ARIMA_Model.pkl"
            saveModel(model,filename)


            def forecast_one_step():
                fc, conf_int = model.predict(n_periods=1, return_conf_int=True)
                return (
                    fc.tolist()[0],
                    np.asarray(conf_int).tolist()[0])
```

5. Run code script ที่เราพัฒนาด้วยคำสั่ง streamlit run

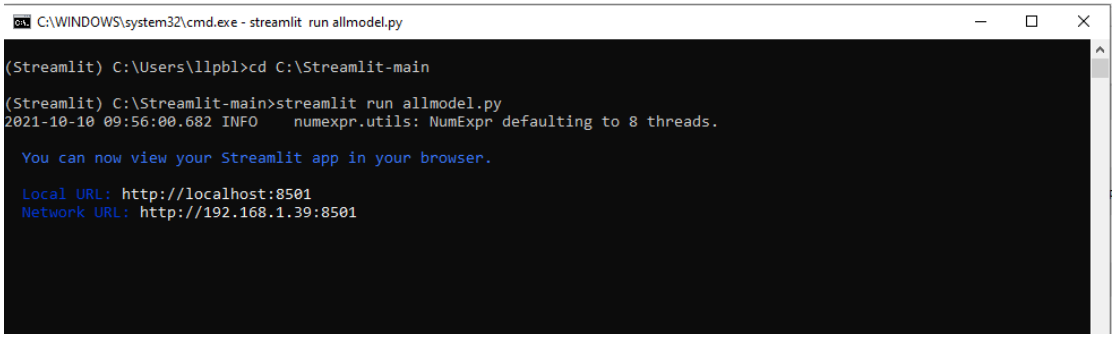
https://docs.streamlit.io/en/stable/main_concepts.html

Create an app

Working with Streamlit is simple. First you sprinkle a few Streamlit commands into a normal Python script, then you run it with `streamlit run`:

```
streamlit run your_script.py [-- script args]
```

As soon as you run the script as shown above, a local Streamlit server will spin up and your app will open in a new tab your default web browser. The app is your canvas, where you'll draw charts, text, widgets, tables, and more.

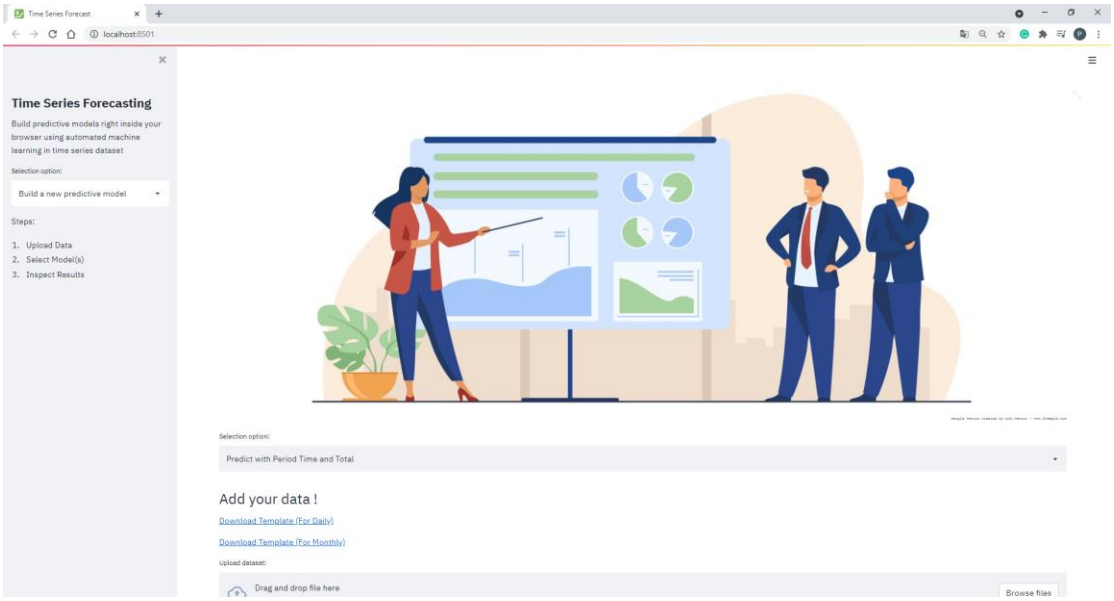


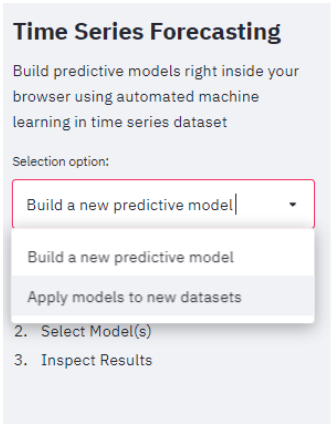
ภาคผนวก ข

ขั้นตอนการใช้งานของระบบวิเคราะห์ข้อมูลอนุกรมเวลา ด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง

ขั้นตอนการใช้งานระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง

1. เข้าสู่หน้าจอของระบบจะมีตัวเลือกด้านซ้าย 2 ส่วน





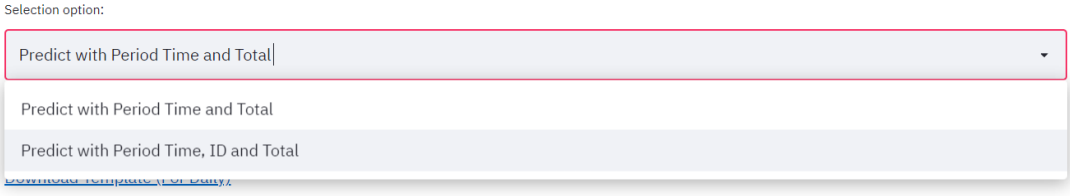
1.1 สร้าง model

1.2 ใช้ model ที่สนใจพยากรณ์ข้อมูลจริง

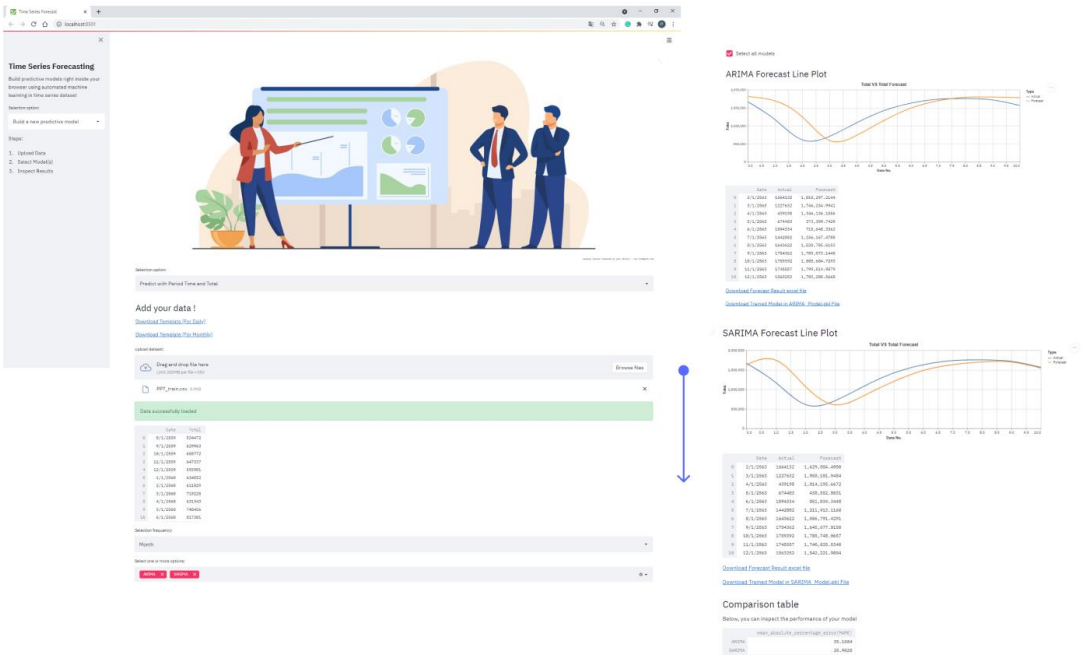
1.1 สร้าง model

เลือกรูปแบบของข้อมูลนำเข้าที่ต้องการพยากรณ์

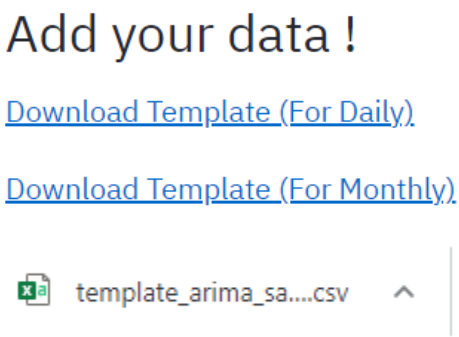
- แบบที่ 1 คือข้อมูล Date, Total
- แบบที่ 2 คือข้อมูล Date, Total, ID



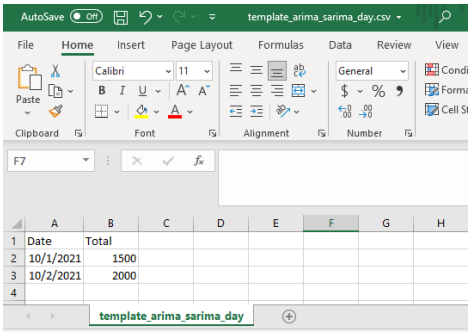
แบบที่ 1 : รูปแบบนี้จะใช้เทคนิค ARIMA, SARIMA สร้าง model



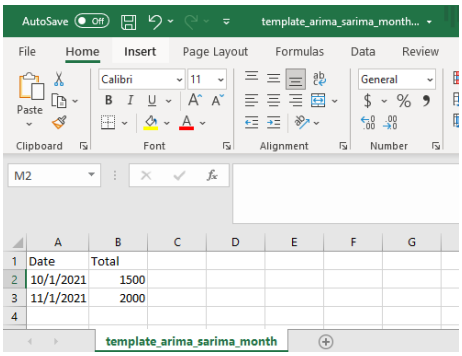
- สามารถ Download Template สำหรับข้อมูลรายวันหรือรายเดือนได้



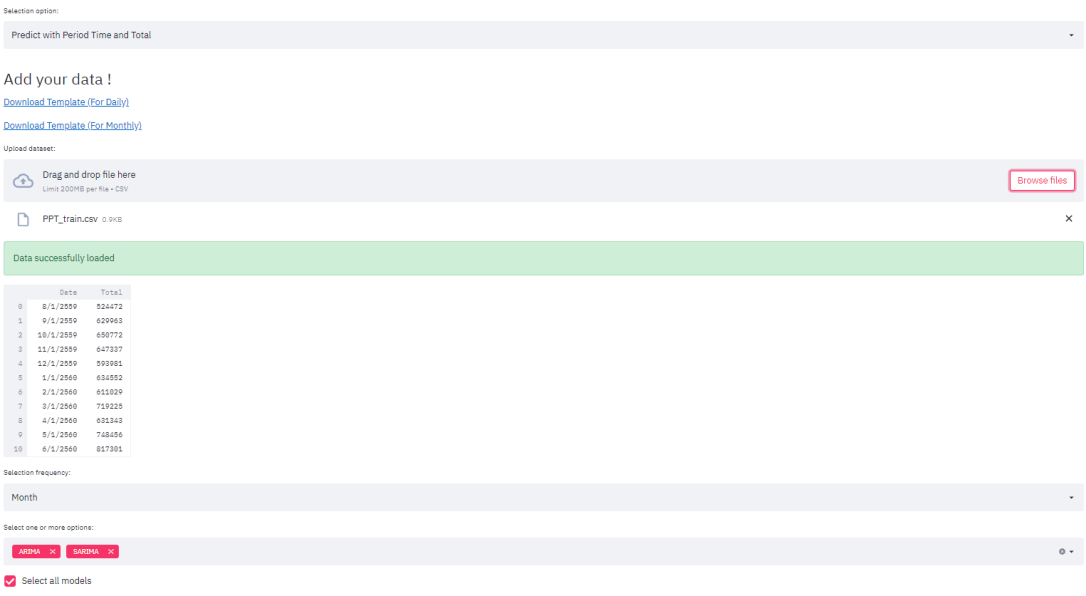
ตัวอย่างไฟล์ template ข้อมูลนำเข้าสร้าง model รายวัน (.csv)



ตัวอย่างไฟล์ template ข้อมูลนำเข้าสร้าง model รายเดือน (.csv)

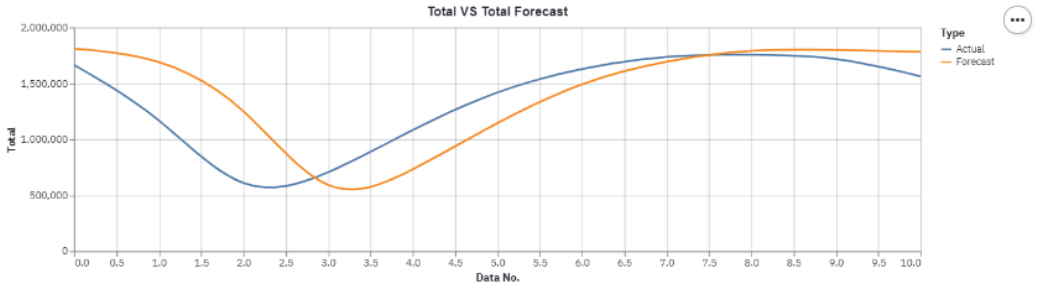


- Upload ข้อมูล .csv เพื่อสร้าง model
- เลือกความช่วงถี่ของข้อมูล (รายวัน, รายเดือน)
- เลือก model ที่ต้องการ หรือเลือกทั้งหมด



- ระบบประมวลผลจากข้อมูลที่ upload เข้าไป 80% train, 20% validate test แสดงผลลัพธ์เปรียบเทียบค่าพยากรณ์และ ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคาดเคลื่อน สามารถ export ค่าข้อมูลเปรียบเทียบในรูปแบบ .xlsx และSave model ที่สนใจได้

ARIMA Forecast Line Plot

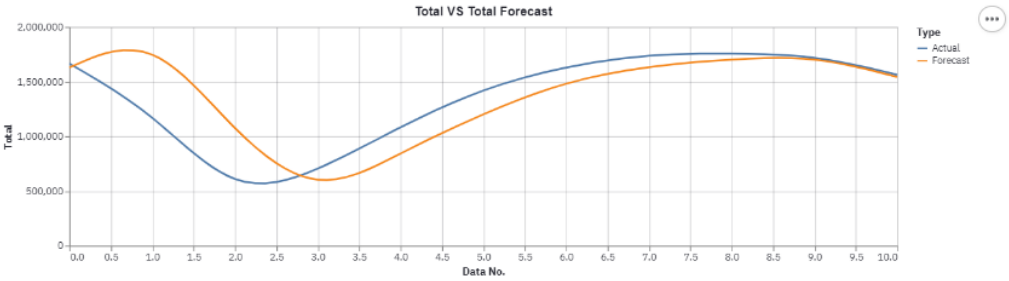


	Date	Actual	Forecast
0	2/1/2563	1664132	1,810,297.2144
1	3/1/2563	1227632	1,746,234.9941
2	4/1/2563	439198	1,346,136.1556
3	5/1/2563	674483	373,399.7429
4	6/1/2563	1094334	715,648.3362
5	7/1/2563	1442882	1,156,167.4705
6	8/1/2563	1643622	1,520,785.6153
7	9/1/2563	1754362	1,709,073.1440
8	10/1/2563	1759392	1,808,604.7293
9	11/1/2563	1745557	1,799,514.9579
10	12/1/2563	1563252	1,783,208.5668

Download Forecast Result excel file

Download Trained Model in ARIMA_Model.pkl File

SARIMA Forecast Line Plot



	Date	Actual	Forecast
0	2/1/2563	1664132	1,629,884.4980
1	3/1/2563	1227632	1,958,101.9484
2	4/1/2563	439198	1,014,195.6672
3	5/1/2563	674483	438,552.8031
4	6/1/2563	1094334	851,034.3448
5	7/1/2563	1442882	1,211,913.1160
6	8/1/2563	1643622	1,506,791.4291
7	9/1/2563	1754362	1,645,677.0150
8	10/1/2563	1759392	1,705,748.0657
9	11/1/2563	1745557	1,740,035.5340
10	12/1/2563	1563252	1,542,221.9004

Download Forecast Result excel file

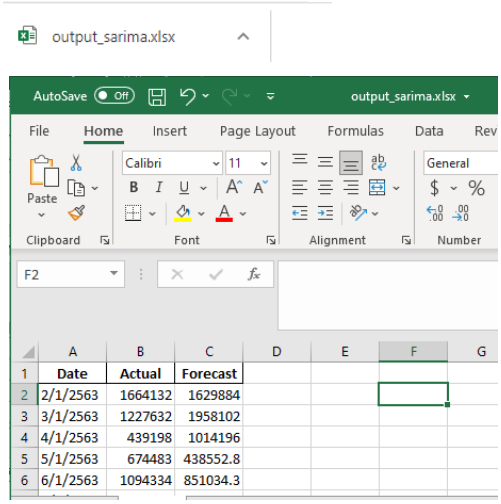
Download Trained Model in SARIMA_Model.pkl File

Comparison table

Below, you can inspect the performance of your model

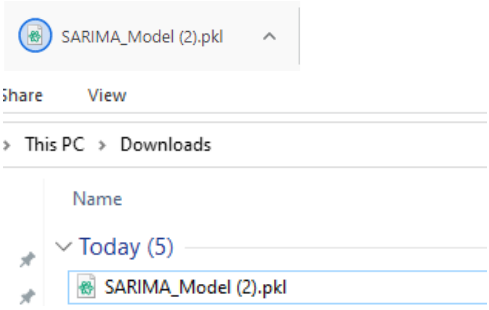
	mean_absolute_percentage_error(MAPE)
ARIMA	35.1504
SARIMA	25.9028

ตัวอย่างไฟล์ข้อมูลเปรียบเทียบในรูปแบบ .xlsx

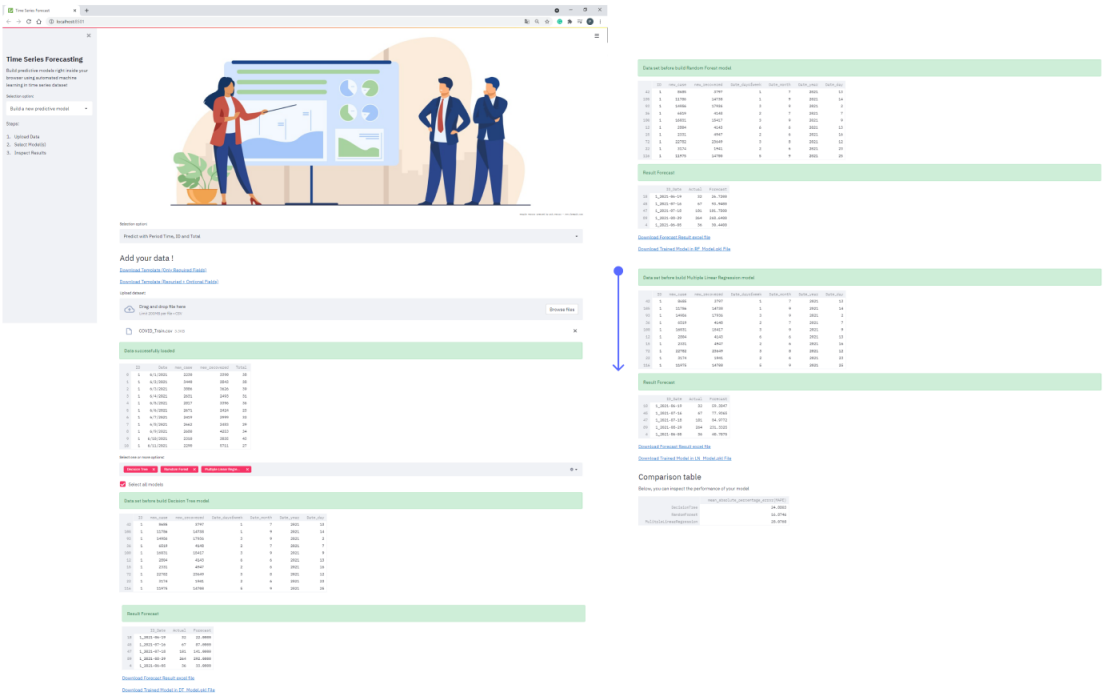


	A	B	C	D	E	F	G
1	Date	Actual	Forecast				
2	2/1/2563	1664132	1629884				
3	3/1/2563	1227632	1958102				
4	4/1/2563	439198	1014196				
5	5/1/2563	674483	438552.8				
6	6/1/2563	1094334	851034.3				

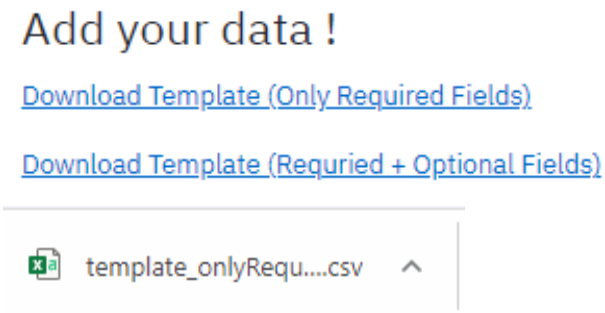
ตัวอย่างไฟล์ model ที่ทำการ download ในรูปแบบ .pkl



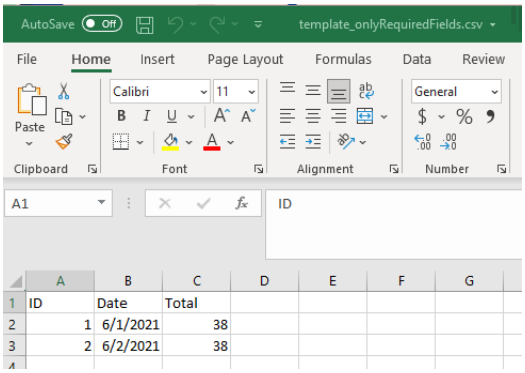
แบบที่ 2 : รูปแบบนี้จะใช้เทคนิค Decision Tree, Random Forest, Multiple Linear Regression สร้าง model



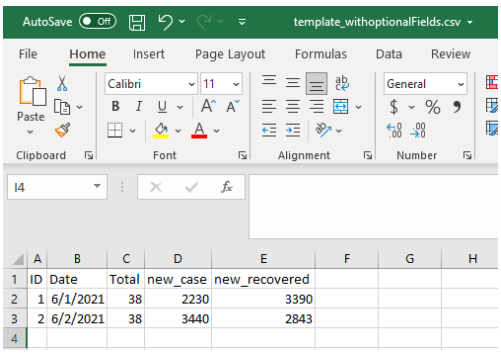
- สามารถ Download Template สำหรับข้อมูลตัวอย่างได้



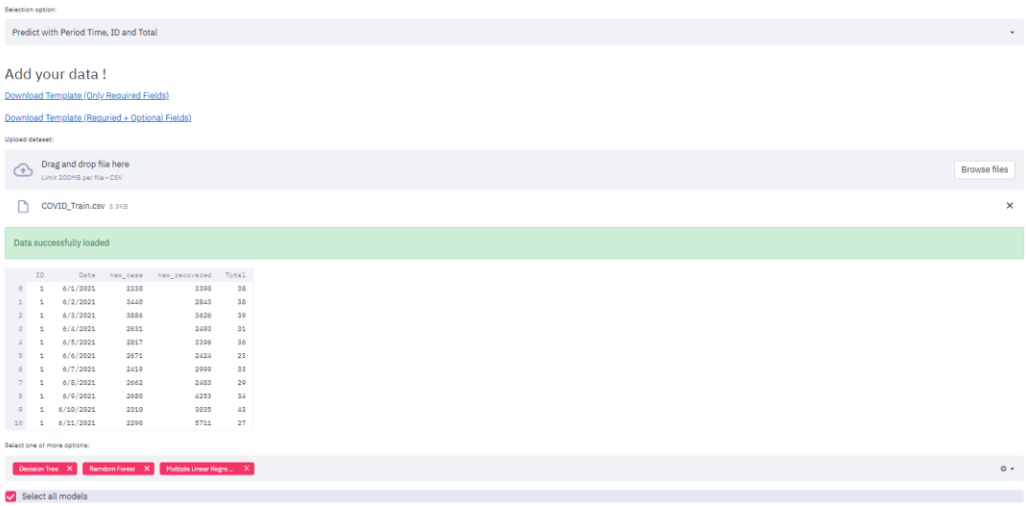
ตัวอย่างไฟล์ template ข้อมูลนำเข้าสร้าง model เฉพาะข้อมูลที่จำเป็นต้องระบุ (.csv) ประกอบไปด้วย ID, Date, Total



ตัวอย่างไฟล์ template ข้อมูลนำเข้าสร้าง model สามารถเพิ่มเติมข้อมูลประเภทตัวเลขเข้าไปเพิ่มเติมได้จากข้อมูลจำเป็นที่ต้องระบุ (ID, Date, Total) (.csv)



- Upload ข้อมูล .csv เพื่อสร้าง model
- เลือก model ที่ต้องการ หรือเลือกทั้งหมด



- ระบบแสดงข้อมูลจริงหลังการแปลงข้อมูลผ่านระบบอัตโนมัติเพื่อสร้าง model

Data set before build Decision Tree model

	ID	new_case	new_recovered	Date_dayofweek	Date_month	Date_year	Date_day
42	1	8685	3797	1	7	2021	13
105	1	11786	14738	1	9	2021	14
93	1	14956	17936	3	9	2021	2
36	1	6519	4148	2	7	2021	7
100	1	16031	15417	3	9	2021	9
12	1	2804	4143	6	6	2021	13
15	1	2331	4947	2	6	2021	16
72	1	22782	23649	3	8	2021	12
22	1	3174	1941	2	6	2021	23
116	1	11975	14700	5	9	2021	25

- ระบบประมวลผลจากข้อมูลที่ upload เข้าไป 80% train, 20% validate test แสดงผลลัพธ์เปรียบเทียบค่าพยากรณ์และ ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคาดเคลื่อน สามารถ export ค่าข้อมูลเปรียบเทียบในรูปแบบ .xlsx และSave model ที่สนใจได้

Result Forecast

	ID_Date	Actual	Forecast
18	1_2021-06-19	32	59.3847
45	1_2021-07-16	67	77.9365
47	1_2021-07-18	101	84.9772
89	1_2021-08-29	264	231.3325
4	1_2021-06-05	36	40.7875

Download Forecast Result excel file

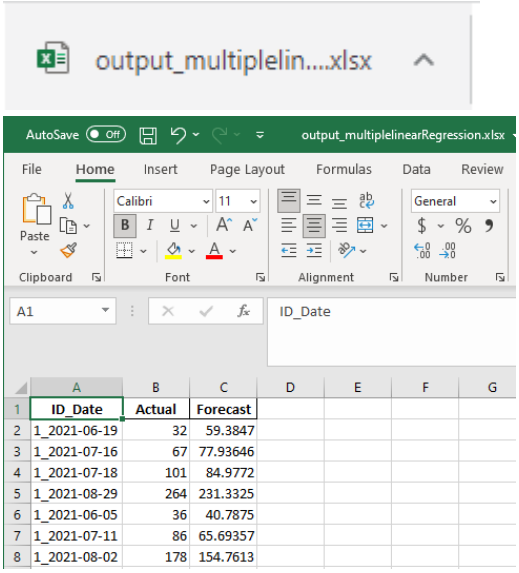
Download Trained Model in LN_Model.pkl File

Comparison table

Below, you can inspect the performance of your model

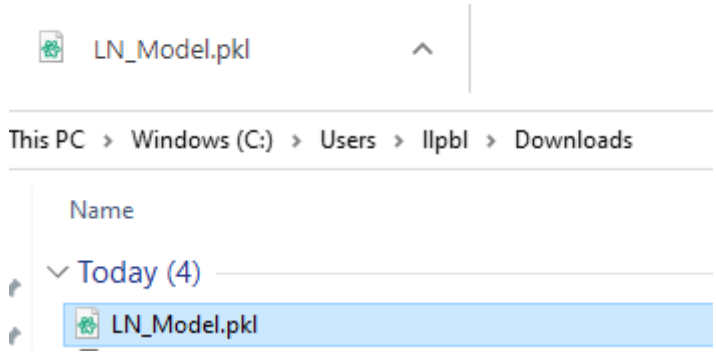
	mean_absolute_percentage_error(MAPE)
DecisionTree	24.8803
RandomForest	16.8746
MulitpleLinearRegression	28.0760

ตัวอย่างไฟล์ข้อมูลเปรียบเทียบในรูปแบบ .xlsx



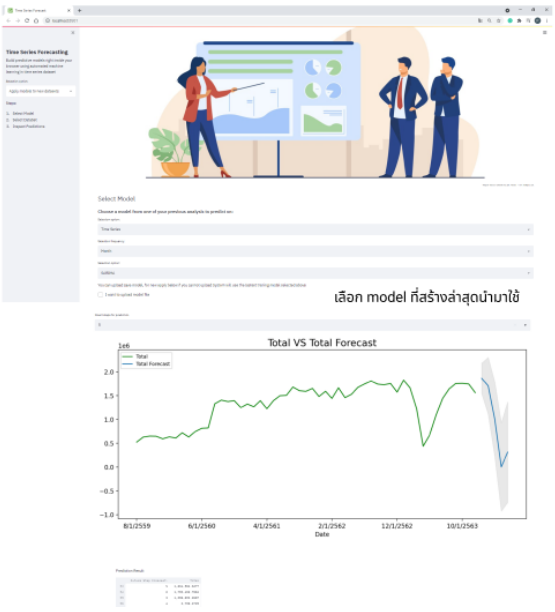
	A	B	C	D	E	F	G
1	ID_Date	Actual	Forecast				
2	1_2021-06-19	32	59.3847				
3	1_2021-07-16	67	77.93646				
4	1_2021-07-18	101	84.9772				
5	1_2021-08-29	264	231.3325				
6	1_2021-06-05	36	40.7875				
7	1_2021-07-11	86	65.69357				
8	1_2021-08-02	178	154.7613				

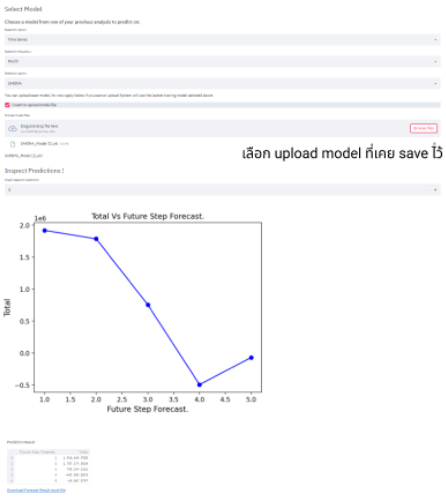
ตัวอย่างไฟล์ model ที่ทำการ download ในรูปแบบ .pkl



1.2 ใช้ model ที่สนใจพยากรณ์ข้อมูลจริง

แบบที่ 1 : รูปแบบนี้จะใช้เทคนิค ARIMA, SARIMA พยากรณ์

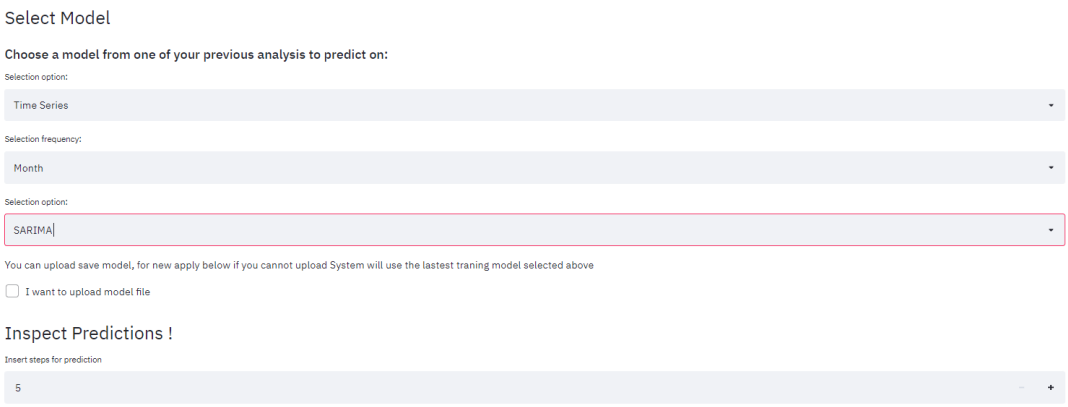




- เลือกประเภทการพยากรณ์ (Time Series)
- เลือกความช่วงถี่ของข้อมูล (รายวัน, รายเดือน)
- เลือก model ที่ต้องการพยากรณ์

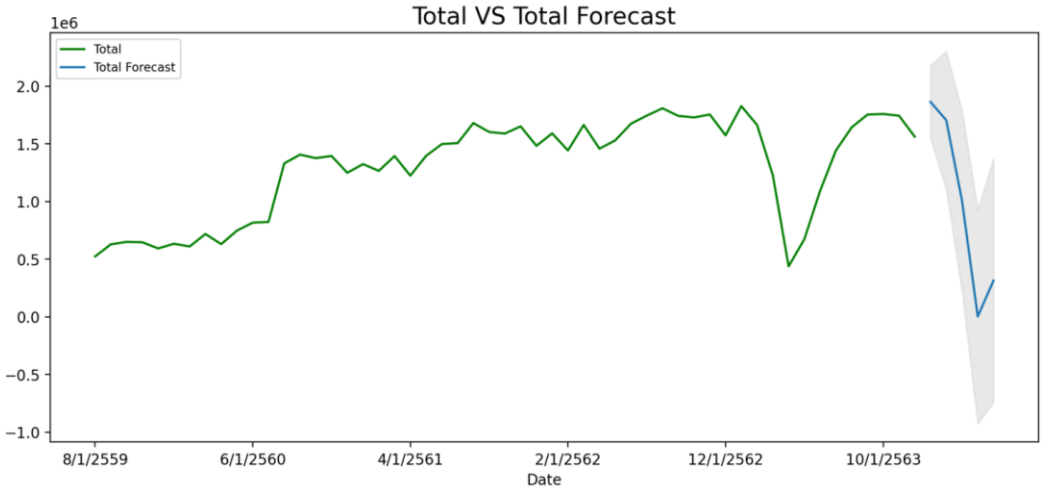
กรณีที่ 1 ระบบจะทำการเก็บ model ล่าสุดที่เพิ่งสร้างไว้หากต้องการเลือกใช้ model ที่เพิ่งสร้างสามารถเลือก model ที่ต้องการพยากรณ์ได้ดังรูปด้านล่าง

ตัวอย่าง กรณีนี้เลือก SARIMA model ช่วงความถี่รายเดือน และต้องการพยากรณ์ล่วงหน้า 5 เดือน



- ระบบประมวลผลและแสดงผลลัพธ์ค่าพยากรณ์

กรณีนี้กราฟจะแสดงข้อมูล train ที่ระบบให้ทำการจัดเก็บไว้ล่าสุดสำหรับสร้าง model (สีเขียว) แสดงเปรียบเทียบกับค่าพยากรณ์ใหม่ด้วย (สีฟ้า)

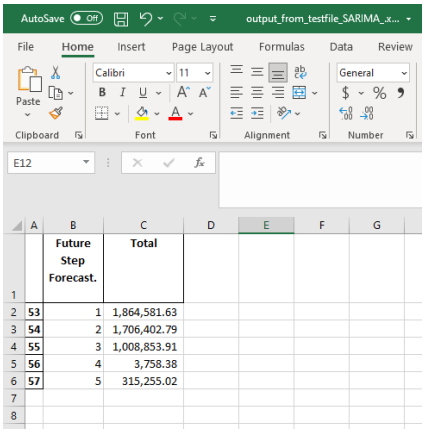


Prediction Result

	Future Step Forecast.	Total
53	1	1,864,581.6277
54	2	1,706,402.7884
55	3	1,008,853.9087
56	4	3,758.3795
57	5	315,255.0196

Download Forecast Result excel file

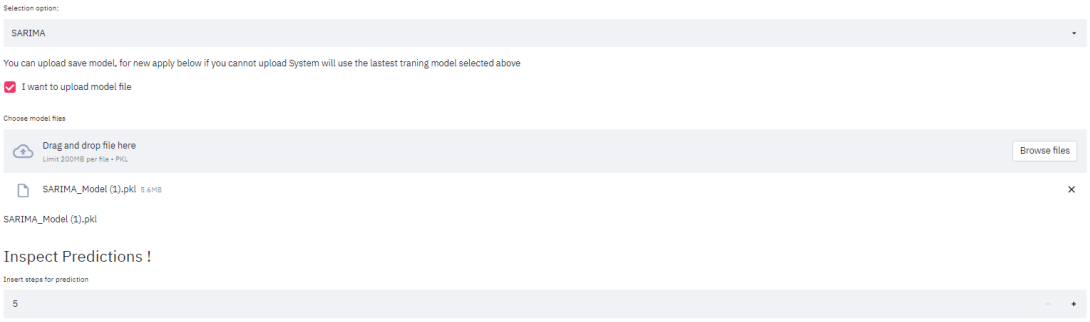
- Export ข้อมูลออกมาในรูปแบบ .xlsx



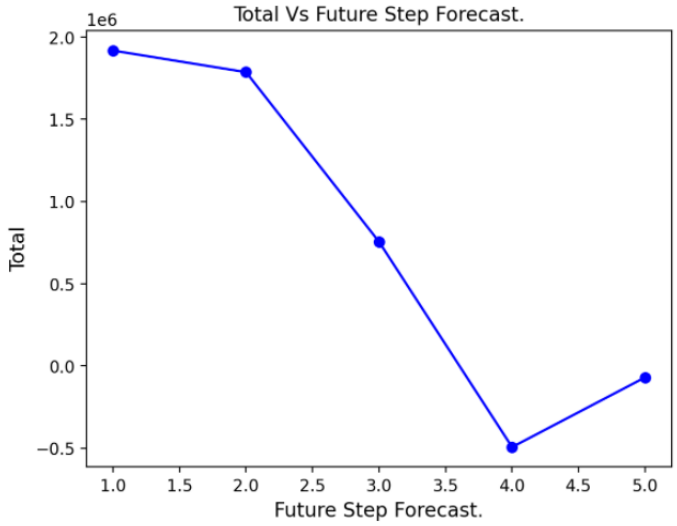
	A	B	C	D	E	F	G
1		Future Step Forecast.	Total				
2	53	1	1,864,581.63				
3	54	2	1,706,402.79				
4	55	3	1,008,853.91				
5	56	4	3,758.38				
6	57	5	315,255.02				
7							
8							

กรณีที่ 2 สามารถเลือก upload model ที่จัดเก็บไว้ (.pkl) สำหรับการพยากรณ์ได้ดังรูปด้านล่าง

ตัวอย่าง กรณีนี้เลือก upload SARIMA model ช่วงความถี่รายเดือน และต้องการพยากรณ์ล่วงหน้า 5 เดือน



- ระบบประมวลผลและแสดงผลลัพธ์ค่าพยากรณ์

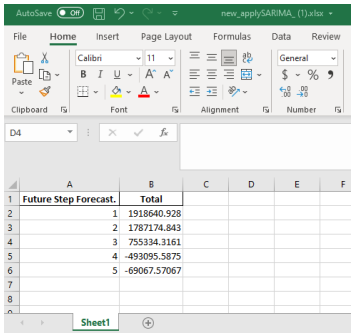


Prediction Result

	Future Step Forecast.	Total
0	1	1,918,640.9283
1	2	1,787,174.8426
2	3	755,334.3161
3	4	-493,095.5875
4	5	-69,067.5707

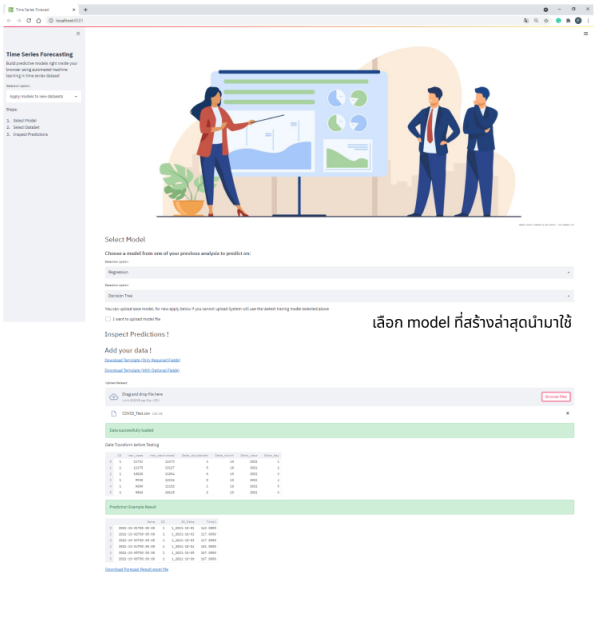
Download Forecast Result excel file

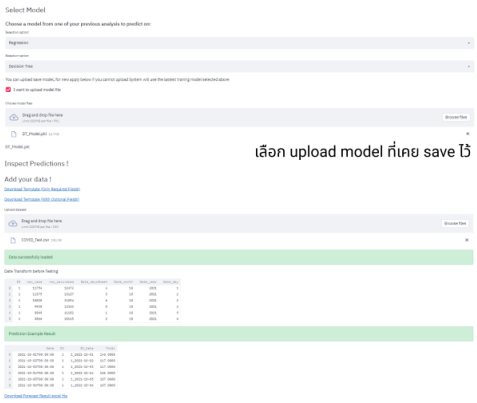
- Export ข้อมูลออกมาในรูปแบบ .xlsx



	A	B	C	D	E	F
1	Future Step Forecast.	Total				
2	1	1918640.928				
3	2	1787174.843				
4	3	755334.3161				
5	4	-493095.5875				
6	5	-69067.57067				
7						
8						

แบบที่ 2 : รูปแบบนี้จะใช้เทคนิค Decision Tree, Random Forest, Multiple Linear Regression พยากรณ์



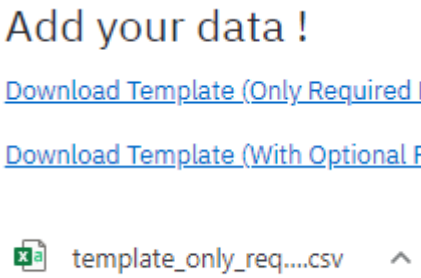


- เลือกประเภทการพยากรณ์ (Regression)
- เลือก model ที่ต้องการพยากรณ์

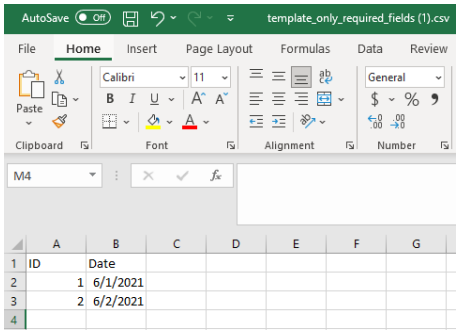
กรณีที่ 1 ระบบจะทำการเก็บ model ล่าสุดที่เพิ่งสร้างไว้หากต้องการเลือกใช้ model ที่เพิ่งสร้างสามารถเลือก model ที่ต้องการพยากรณ์ได้ดังรูปด้านล่าง

ตัวอย่าง กรณีนี้เลือก Decision tree model

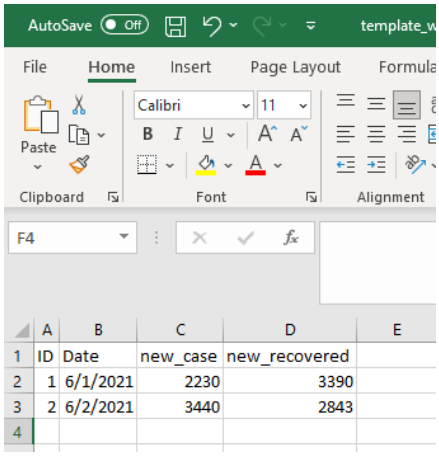
• สามารถ Download Template สำหรับข้อมูลตัวอย่างได้



ตัวอย่างไฟล์ template ข้อมูลสำหรับพยากรณ์เฉพาะข้อมูลที่จำเป็นต้องระบุ (.csv) ประกอบไปด้วย ID, Date

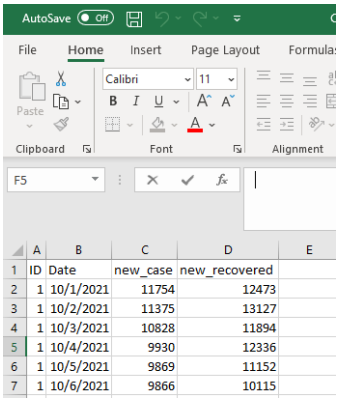


ตัวอย่างไฟล์ template ข้อมูลสำหรับพยากรณ์สามารถเพิ่มเติมข้อมูลประเภทตัวเลขเข้าไปเพิ่มเติมได้จากข้อมูลจำเป็นที่ต้องระบุ (ID, Date) (.csv)



- Upload ข้อมูล .csv ที่ต้องการให้พยากรณ์ในอนาคต

  ตัวอย่าง : กรณีต้องการพยากรณ์จำนวนผู้เสียชีวิตรายใหม่จากสถานการณ์ COVID-19 ระลอก 3 วันที่1-6 ตุลาคม 2564



	A	B	C	D	E
1	ID	Date	new_case	new_recovered	
2	1	10/1/2021	11754	12473	
3	1	10/2/2021	11375	13127	
4	1	10/3/2021	10828	11894	
5	1	10/4/2021	9930	12336	
6	1	10/5/2021	9869	11152	
7	1	10/6/2021	9866	10115	

- ระบบแสดงข้อมูลจริงหลังการแปลงข้อมูลผ่านระบบอัตโนมัติเพื่อนำเข้าพยากรณ์

Date Transform before Testing

	ID	new_case	new_recovered	Date_dayofweek	Date_month	Date_year	Date_day
0	1	11754	12473	4	10	2021	1
1	1	11375	13127	5	10	2021	2
2	1	10828	11894	6	10	2021	3
3	1	9930	12336	0	10	2021	4
4	1	9869	11152	1	10	2021	5
5	1	9866	10115	2	10	2021	6

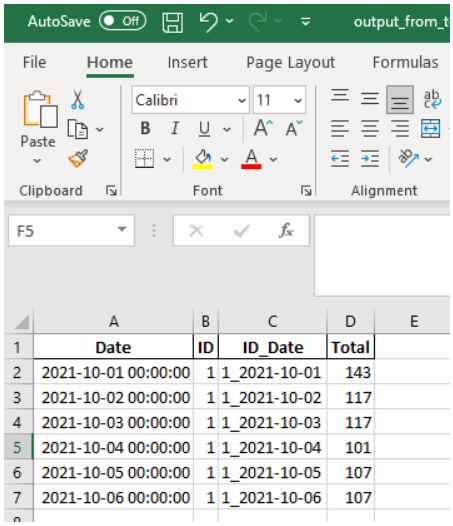
- ระบบประมวลผลและแสดงผลลัพธ์ค่าพยากรณ์

Prediction Example Result

	Date	ID	ID_Date	Total
0	2021-10-01T00:00:00	1	1_2021-10-01	143.0000
1	2021-10-02T00:00:00	1	1_2021-10-02	117.0000
2	2021-10-03T00:00:00	1	1_2021-10-03	117.0000
3	2021-10-04T00:00:00	1	1_2021-10-04	101.0000
4	2021-10-05T00:00:00	1	1_2021-10-05	107.0000
5	2021-10-06T00:00:00	1	1_2021-10-06	107.0000

Download Forecast Result excel file

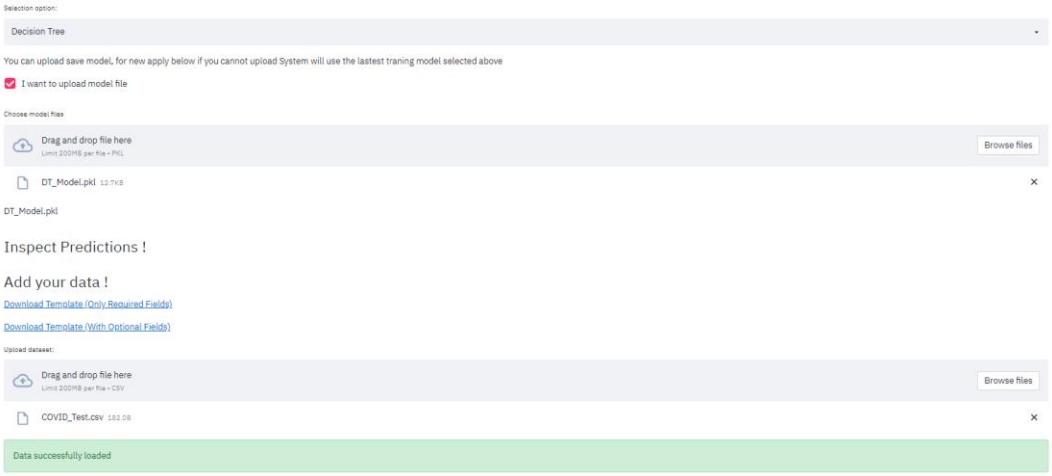
- Export ข้อมูลออกมาในรูปแบบ .xlsx



	A	B	C	D	E
1	Date	ID	ID_Date	Total	
2	2021-10-01 00:00:00	1	1_2021-10-01	143	
3	2021-10-02 00:00:00	1	1_2021-10-02	117	
4	2021-10-03 00:00:00	1	1_2021-10-03	117	
5	2021-10-04 00:00:00	1	1_2021-10-04	101	
6	2021-10-05 00:00:00	1	1_2021-10-05	107	
7	2021-10-06 00:00:00	1	1_2021-10-06	107	

กรณีที่ 2 สามารถเลือก upload model ที่จัดเก็บไว้ (.pkl) สำหรับการพยากรณ์ได้ดังรูปด้านล่างตัวอย่าง

กรณีนี้เลือก Upload Decision tree model และ upload ข้อมูลสำหรับพยาการณ์จำนวนผู้เสียชีวิตรายใหม่จากสถานการณ์ COVID-19 ระลอก 3 วันที่1-6 ตุลาคม 2564



- ระบบแสดงข้อมูลจริงหลังการแปลงข้อมูลผ่านระบบอัตโนมัติเพื่อนำเข้าพยากรณ์

Date Transform before Testing

	ID	new_case	new_recovered	Date_dayofweek	Date_month	Date_year	Date_day
0	1	11754	12473	4	10	2021	1
1	1	11375	13127	5	10	2021	2
2	1	10828	11894	6	10	2021	3
3	1	9930	12336	0	10	2021	4
4	1	9869	11152	1	10	2021	5
5	1	9866	10115	2	10	2021	6

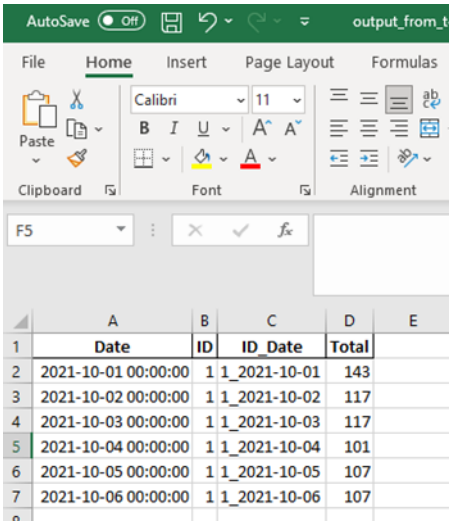
- ระบบประมวลผลและแสดงผลลัพธ์ค่าพยากรณ์

Prediction Example Result

	Date	ID	ID_Date	Total
0	2021-10-01T00:00:00	1	1_2021-10-01	143.0000
1	2021-10-02T00:00:00	1	1_2021-10-02	117.0000
2	2021-10-03T00:00:00	1	1_2021-10-03	117.0000
3	2021-10-04T00:00:00	1	1_2021-10-04	101.0000
4	2021-10-05T00:00:00	1	1_2021-10-05	107.0000
5	2021-10-06T00:00:00	1	1_2021-10-06	107.0000

Download Forecast Result excel file

Export ข้อมูลออกมาในรูปแบบ .xlsx



ประวัติผู้เขียน

ชื่อ-นามสกุล	นางสาว พรทิวา วิศิษฏ์สรอรรถ
ประวัติการศึกษา	วิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ปีการศึกษา 2556
ตำแหน่งและสถานที่ทำงานปัจจุบัน	ผู้จัดการ ฝ่ายพัฒนาผลิตภัณฑ์และการใช้ดิจิทัลเพื่อลูกค้าธุรกิจ SME สายงานพัฒนาธุรกิจดิจิทัลโซลูชั่นเพื่อลูกค้า ธุรกิจ กลุ่มงานนวัตกรรมดิจิทัลและข้อมูล ธนาคารกรุงศรีอยุธยา จำกัด (มหาชน)