

การพยากรณ์ปริมาณการเข้าชมเว็บไซต์สำหรับธุรกิจ อีคอมเมิร์ซโดยใช้เทคนิคการเรียนรู้ของเครื่องและสร้างระบบ แจ้งเตือนโดยอัตโนมัติ

Forecasting web traffic for e-commerce website using
machine learning and automated push notifications

รศ.ดร. วุฒิชัย ร่มสายหยุด
ประธานกรรมการ

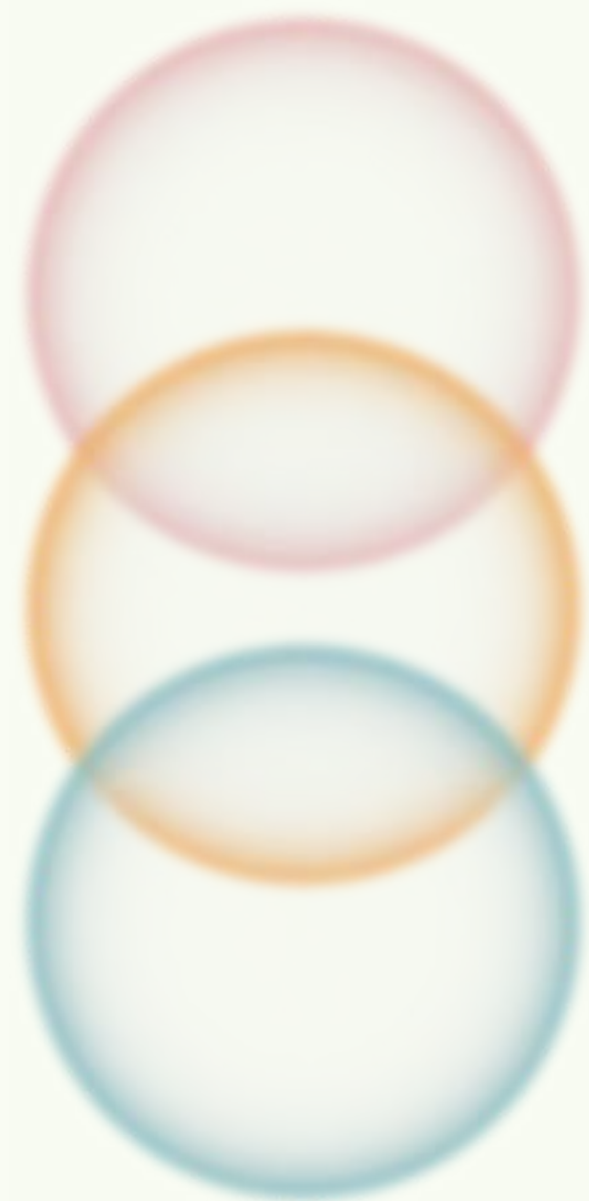
ผศ.ดร. ดวงใจ จิตคงชื่น
กรรมการ

ดร. เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา
อาจารย์ที่ปรึกษา

สิทธี โฉ่ววัฒนา
645162020022
ผู้จัดทำ

Agenda

- ที่มาและความสำคัญ
- วัตถุประสงค์และขอบเขตงาน
- แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง
- ระเบียบวิธีวิจัย
- ผลการดำเนินงาน
- สรุปผลและข้อเสนอแนะ



ที่มาและความสำคัญ

ที่มาและความสำคัญ

- เคยประสบปัญหา Regional Team มาขอเยี่ยมทรัพยากรไปถึง 30% เพื่อที่จะนำไปช่วยในส่วนของ marketplace ทำให้ทั้งแผนกทำงานได้ช้าลง เนื่องจากทรัพยากรในการประมวลผลไม่เพียงพอ
- ความคิดเห็นจากผู้ใช้งานในทางลบจาก Social Media ที่เกิดจากปัญหาเว็บล่มหรือแอปพลิเคชันใช้งานไม่ได้
- ผู้ใช้งานย้ายไปใช้งานบนแพลตฟอร์มของกลุ่มแข่งขัน เนื่องจากมีประสบการณ์การใช้งานที่ไม่ดี

ตัวอย่างความคิดเห็นของผู้ใช้งานจาก Social Media



VI * IV 🧡 Move Further Get more 🧡 @LoveRealmTh · Jul 7
App shopee ล่ม



86



ข้อป้ ล่มออทำไมให้ลื้อคอื่นใหม่ ?

#ShopeeTH #shopee



126



977



403



212.9K



Stone.rose @Stonero18717930 · Apr 4
lazada คือไรอะ ล่มเฉย #lazadaล่ม



1



1



865



น้ำใจ (-_-) @naamjai · Apr 4
Lazada ล่ม?



537



LATTTE : ฤเหมียว @chaerinnuna · Apr 4

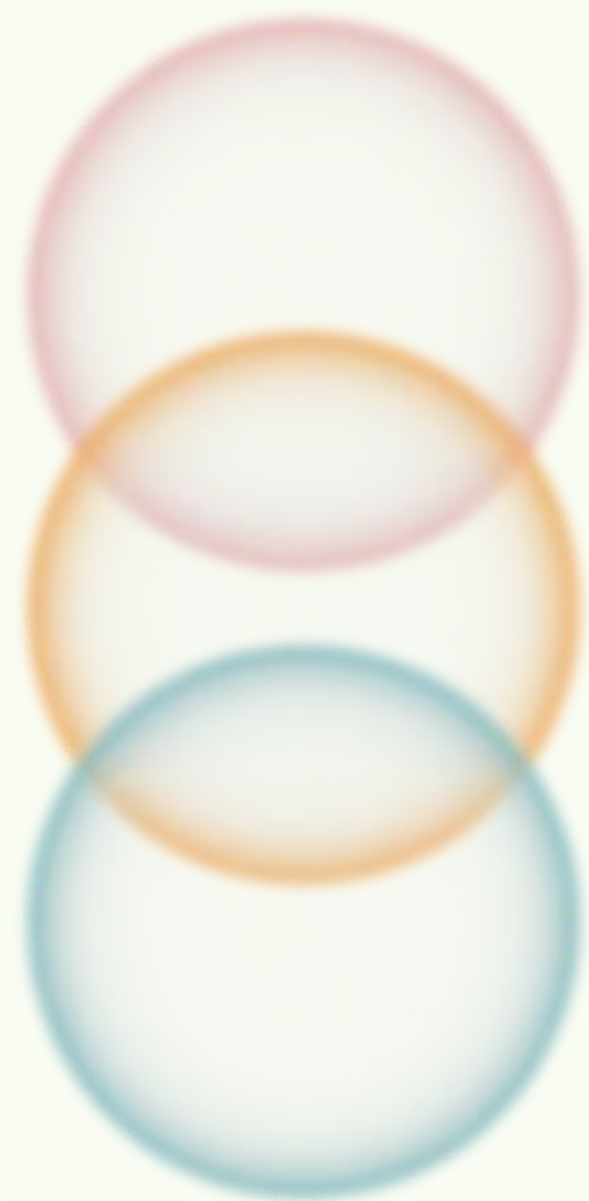
ลาซาด้า ล่มเหรวะะะะะะะ



2



519



วัตถุประสงค์และ ขอบเขตงาน

วัตถุประสงค์

- เพื่อกำหนดปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วัน
- นำผลที่ได้จากการทำนายมาใช้ในการแจ้งเตือนผ่านแอปพลิเคชัน เพื่อที่จะได้มีเวลาเตรียมการรับมือกับปัญหา
- ศึกษาพฤติกรรมการใช้งานของผู้ใช้เพื่อประโยชน์อื่นๆทางธุรกิจ
- เสนอและเปรียบเทียบแบบจำลองต่างๆ เพื่อเลือกใช้ให้เหมาะสมกับประเภทของงาน

ขอบเขตงาน

- ใช้การดึงข้อมูล Web Traffic จาก Google Analytics
- ข้อมูลรายวันของ Google Merchandise Store ระหว่างวันที่ 1 มกราคม 2559 ถึง 6 พฤษภาคม 2566
- ข้อมูลแบบ Structured Data



แนวคิด ทฤษฎีและ งานวิจัยที่เกี่ยวข้อง

แนวคิดและทฤษฎี

- ความหมายและประเภทของ Web Traffic
- ความหมายและประเภทของการพยากรณ์
- พยากรณ์แบบอนุกรมเวลา (Time Series Forecasting)
- Autoregressive Integrated Moving Average (ARIMA)
- Prophet

ความหมายของ Web Traffic

จำนวนข้อมูลที่ถูกส่งและรับผ่านเครือข่ายอินเทอร์เน็ตของเว็บไซต์หรือแอปพลิเคชันเว็บในระยะเวลาที่กำหนด โดยมักใช้เพื่อวัดปริมาณการเข้าชมหรือการใช้งานในแต่ละวัน จำนวน web traffic สามารถสื่อถึงความนิยมและความสำคัญของเว็บไซต์หรือแอปพลิเคชันเว็บได้ในบางครั้ง ยังมีจำนวน web traffic มากขึ้น มักแปลว่ามีผู้เข้าชมหรือผู้ใช้งานมากขึ้นเช่นกัน

ประเภทของ Web Traffic

- **Organic Traffic**

เป็นการเข้าชมเว็บไซต์หรือแอปพลิเคชันเว็บที่เกิดจากผลการค้นหาธรรมชาติในเครื่องมือค้นหา เช่น ผลการค้นหาใน Google ผู้เข้าชมประเภทนี้มักเป็นผู้ที่มีความสนใจในเนื้อหาที่เว็บไซต์หรือแอปพลิเคชันเว็บมีอยู่แล้ว

- **Referral Traffic**

เป็นการเข้าชมที่เกิดจากการคลิกลิงก์ที่อยู่ในเว็บไซต์อื่นๆ ที่ชี้ไปยังเว็บไซต์หรือแอปพลิเคชันเว็บปลายทาง เช่น ลิงก์ที่อยู่ในโพสต์บนสื่อสังคมออนไลน์หรือลิงก์ที่อยู่ในบทความบนเว็บไซต์อื่น

ประเภทของ Web Traffic

- **Direct Traffic**

เป็นการเข้าชมที่เกิดจากผู้ใช้งานที่พิมพ์ URL ของเว็บไซต์หรือแอปพลิเคชันเว็บโดยตรงในแถบที่อยู่ของเบราว์เซอร์ หรือใช้บุ๊กมาร์กหรือประวัติการเรียกดูเว็บ ไม่ได้เกิดจากการค้นหาบน Search Engine

- **Social Media Traffic**

เป็นการเข้าชมที่เกิดจากการคลิกลิงก์ที่อยู่ในโพสต์หรืออัปเดตทางสื่อสังคมออนไลน์ เช่น Facebook, Twitter, Instagram เป็นต้น

ประเภทของ Web Traffic

- **Paid Traffic**

เป็นการเข้าชมที่เกิดจากการโฆษณาและการตลาดทางออนไลน์ที่เสียค่าใช้จ่าย เช่น การโฆษณาบนเครื่องมือค้นหาหรือโฆษณาบนโซเชียลมีเดีย

- **Campaign Traffic**

เป็นการเข้าชมที่เกิดจากแคมเปญการตลาดที่เฉพาะเจาะจง เช่น การส่งอีเมลโปรโมชันหรือการโฆษณาทางอีเมล

ความหมายของการพยากรณ์

การพยากรณ์เป็นกระบวนการที่ใช้ข้อมูลที่มีอยู่ในอดีตหรือปัจจุบันเพื่อทำนายหรือคาดคะเนสิ่งที่จะเกิดขึ้นในอนาคต ซึ่งสามารถนำไปใช้ในการวิเคราะห์ให้เห็นแนวโน้มของการเปลี่ยนแปลงในสิ่งที่เราต้องการพยากรณ์ โดยใช้การวิเคราะห์และการสร้างแบบจำลองทางสถิติและคณิตศาสตร์ เพื่อประมาณค่าหรือเหตุการณ์ที่อาจเกิดขึ้นในอนาคต

ประเภทของการพยากรณ์

ประเภทของการพยากรณ์ตามระยะเวลา

- การพยากรณ์ระยะสั้น (Short-term Forecasting) เป็นการพยากรณ์ในช่วงเวลาไม่เกิน 1 ปี
- การพยากรณ์ระยะปานกลาง (Intermediate Forecasting) เป็นการพยากรณ์ในช่วงเวลาระหว่าง 3 เดือนถึง 3 ปี
- การพยากรณ์ระยะยาว (Long-term Forecasting) เป็นการพยากรณ์ในช่วงเวลามากกว่า 3 ปีขึ้นไป

ประเภทของการพยากรณ์

ประเภทของการพยากรณ์ตามคุณลักษณะ

- วิธีการเชิงคุณภาพ (Qualitative Methods) เป็นการพยากรณ์โดยอาศัยผู้เชี่ยวชาญในการคาดการณ์ระยะสั้น โดยไม่ใช้แบบจำลองทางคณิตศาสตร์ จึงตรวจสอบความแม่นยำได้ยากกว่าการพยากรณ์เชิงปริมาณ
- วิธีการเชิงปริมาณ (Quantitative Methods) จะใช้สถิติที่มีความหมายและข้อมูลเก่าเพื่อคาดการณ์แนวโน้มในอนาคตในระยะยาว

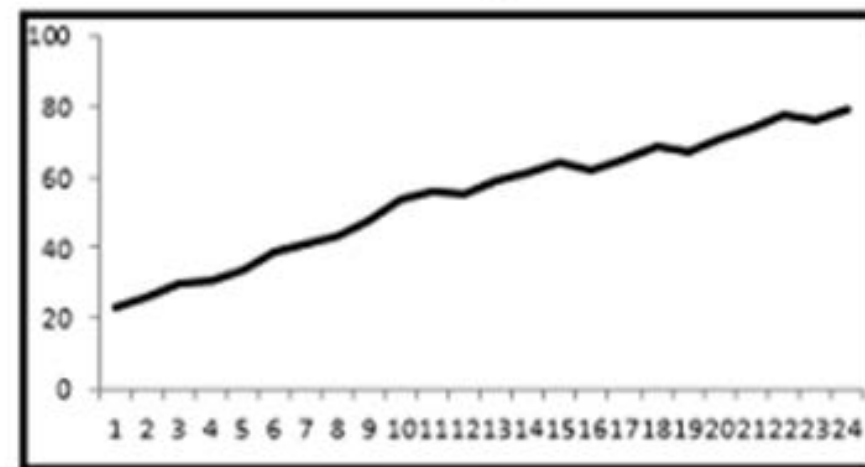
Time Series Forecasting

เป็นวิธีการอาศัยข้อมูลจากอดีตเพื่อพยากรณ์หรือคาดคะเนสิ่งที่เกิดขึ้นในอนาคต องค์ประกอบของอนุกรมเวลา ประกอบไปด้วย 4 ส่วนดังนี้

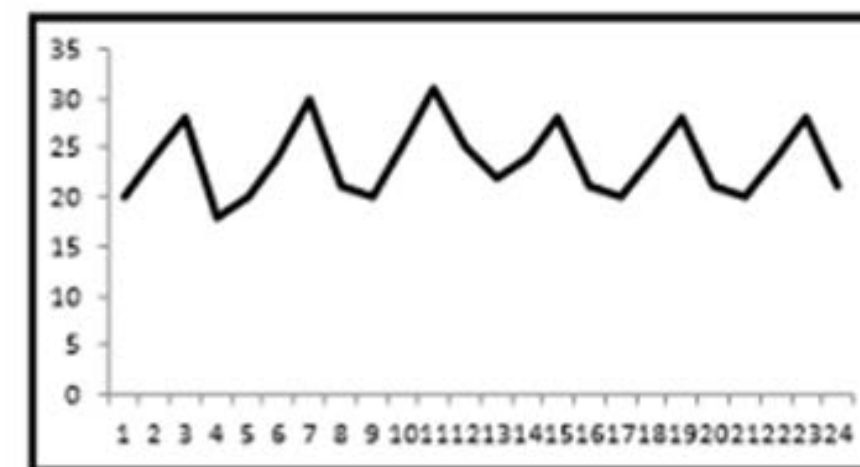
- แนวโน้ม (Trend) มีรูปแบบการเคลื่อนไหวซ้ำๆ ของการเพิ่มขึ้นหรือลดลงอย่างค่อยเป็นค่อยไป
- ฤดูกาล (Season) มีรูปแบบการเปลี่ยนแปลงแบบซ้ำๆ กัน จากผลกระทบของฤดูกาล
- วัฏจักร (Cycle) มีรูปแบบการเคลื่อนไหวซ้ำๆ ของการเพิ่มขึ้นหรือลดลง จะแสดงรูปแบบให้เห็น โดยไม่เกี่ยวข้องกับฤดูกาล สามารถมีระยะเวลายาวเท่าหรือยาวกว่าฤดูกาล
- ความผิดปกติหรือความไม่แน่นอน (Irregular or Random) มีรูปแบบที่ไม่มีแบบแผนเป็นผลกระทบจากการเปลี่ยนแปลงแบบสุ่ม

องค์ประกอบทั้งสี่ของ Time Series

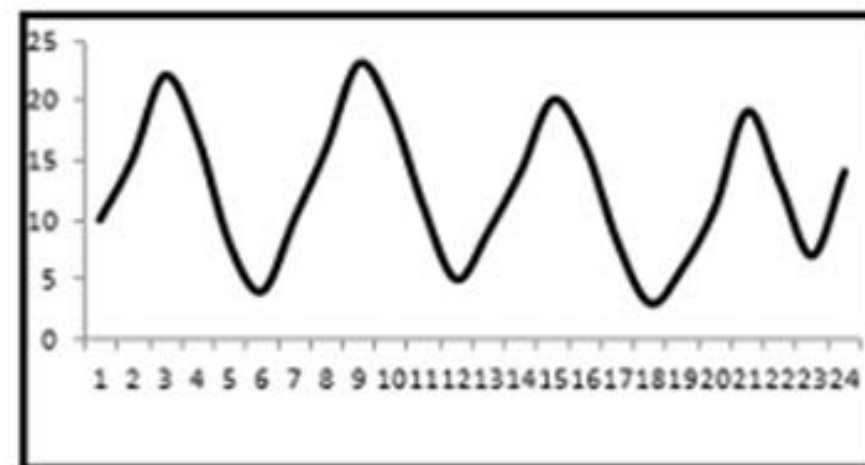
Components of Time Series



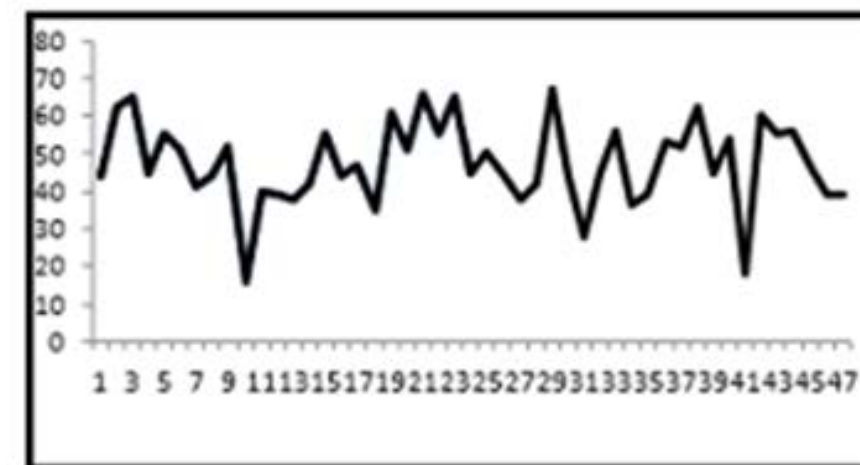
(a) Trend



(b) Seasonality



(c) Cyclicality



(d) Irregular

ARIMA

เป็นแบบจำลองเชิงสถิติที่ใช้ในการพยากรณ์และการวิเคราะห์อนุกรมเวลา (Time Series Analysis) โดยเฉพาะ โดยมีความสามารถในการจัดการกับองค์ประกอบของแนวโน้ม (Trend) และสภาวะซ้ำซ้อน (Seasonality) ซึ่งเป็นความสามารถที่สำคัญในการพยากรณ์และการวิเคราะห์แนวโน้มของข้อมูลตามเวลาในอนาคต

ARIMA ประกอบด้วยสามส่วนหลัก คือ AR (Autoregressive), I (Integrated), และ MA (Moving Average)

Prophet

PROPHET

เป็นไลบรารีที่สามารถใช้งานได้ทั้งใน R และ Python ซึ่งช่วยให้ผู้ใช้วิเคราะห์และทำนายค่าชุดข้อมูลแบบ Time Series ได้อย่างง่ายดาย ไลบรารีนี้ปล่อยตัวออกมาในปี 2017 ด้วยความพยายามของนักพัฒนาที่ให้การวิเคราะห์ข้อมูลแบบ Time Series ให้สามารถใช้งานได้โดยไม่ต้องมีความเชี่ยวชาญมากนัก มีความเป็นมิตรต่อผู้ใช้สูง แต่ยังคงสามารถปรับแต่งได้มาก

Prophet

Prophet สร้างแบบจำลองสำหรับพยากรณ์ชุดข้อมูลเวลาที่
สามารถแยกส่วนประกอบหลัก ได้แก่ แนวโน้ม ฤดูกาล วันหยุด และความ
ผิดปกติ ดังสมการต่อไปนี้

$$y(t) = g(t) + h(t) + s(t) + et$$

Additive Regressive Model

Trend Factor

Holiday component

Seasonality Component

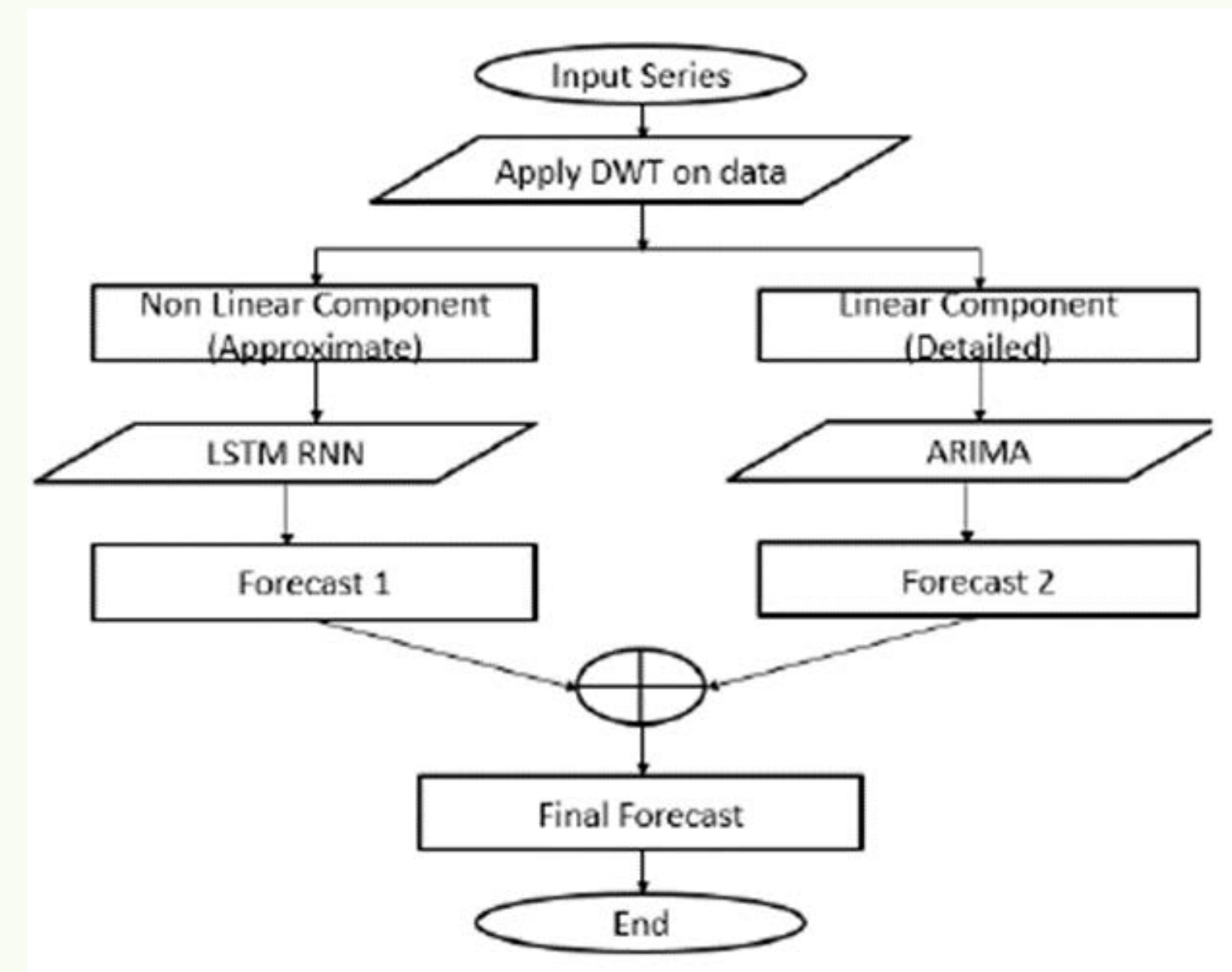
Error term

งานวิจัยที่เกี่ยวข้อง

Web Traffic Time Series Forecasting using ARIMA and LSTM RNN

Shelatkar, T., Tondale, S., Yadav, S., & Ahir, S.
(2020)

ได้นำเสนอการพยากรณ์ข้อมูลอนุกรมเวลาด้วยการ ใช้ Discrete Wavelet Transform (DWT) แบ่งข้อมูลที่ได้มาจาก Wikipedia Pageview API ออกเป็น 2 ส่วน ในขั้นตอนการจัดเตรียมข้อมูล ทำให้ได้ส่วนที่เป็น Detail และ Approximate เพื่อที่จะแบ่งให้ ARIMA และ LSTM RNN ช่วยกันพยากรณ์ และจะนำผลการพยากรณ์กลับมา รวมกันด้วยวิธี Invert Discrete Wavelet Transform (iDWT) ทำให้ได้ผลการพยากรณ์ที่แม่นยำมากขึ้น



งานวิจัยที่เกี่ยวข้อง

Website Traffic Time Series Forecasting Using Regression Machine Learning

D. Sikka and C. N. S. Vinoth Kumar (2023)

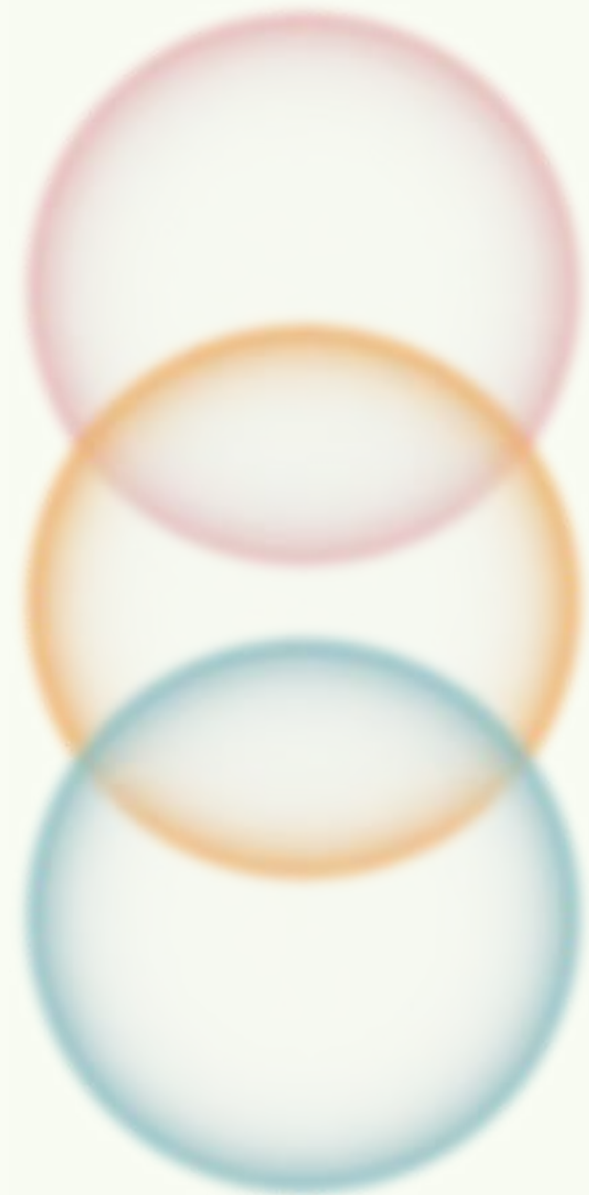
ได้นำเสนอการพยากรณ์ข้อมูลอนุกรมเวลาด้วยการใช้ Ensemble model ที่ประกอบไปด้วยแบบจำลองหลายแบบได้แก่ Decision Tree, Multiple Linear Regression และ Support Vector Machine ในการพยากรณ์ข้อมูล pageview และ visitors ในการหาจำนวน unique visitors จาก statforecasting.com

งานวิจัยที่เกี่ยวข้อง

Web Traffic Prediction of Wikipedia Pages

N. Petluri and E. Al-Masri (2018)

ได้นำเสนอการพยากรณ์อนุกรมเวลาด้วยวิธีการสร้างตัวแปรใหม่ (Feature Engineering) จากแบบจำลองเดิมที่มีคือ RNN seq2seq โดยที่ผู้วิจัยได้ตั้งสมมติฐานขึ้นมาว่าการสร้างตัวแปรใหม่ขึ้นมา จะช่วยส่งเสริมการเรียนรู้ให้กับแบบจำลอง ซึ่งจะทำให้ประสิทธิภาพของแบบจำลองที่มีอยู่นั้นสูงขึ้นได้



ระเบียบวิธีวิจัย

ขั้นตอนการดำเนินงาน

การรวบรวมข้อมูล

- Google Analytics

การจัดเตรียมข้อมูล

- เปลี่ยนรูปแบบข้อมูล
- ตั้งชื่อคอลัมน์
- ลบข้อมูลที่ไม่เกี่ยวข้อง
- จัดเรียงข้อมูล
- สร้างตัวแปรใหม่
(Feature Engineering)

การสร้างแบบจำลอง

- สร้างแบบจำลอง
 - ARIMA
 - Prophet (Uni.)
 - Prophet (Multi.)

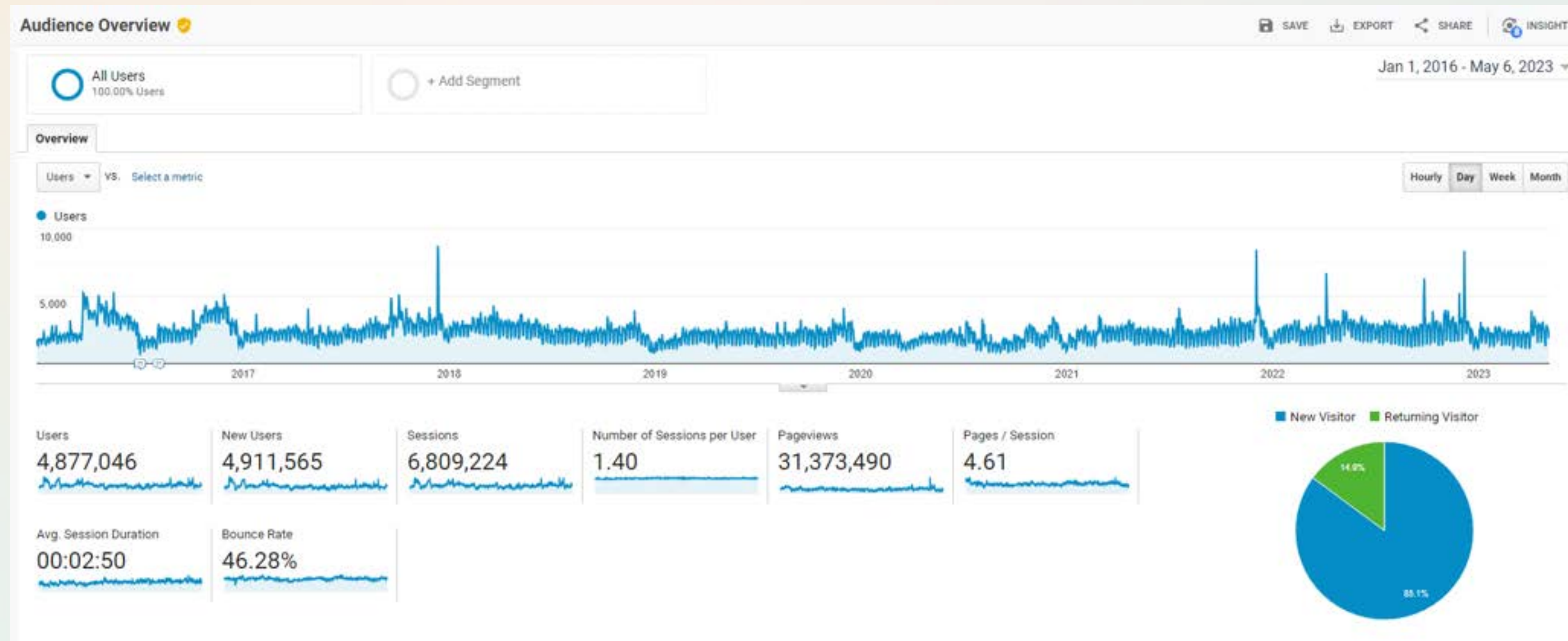
การนำไปใช้งาน

- สร้างระบบแจ้งเตือนผ่าน
แอปพลิเคชัน

การรวบรวมข้อมูล

- รวบรวมข้อมูล Google Merchandise Store จาก Google Analytics
- ข้อมูลระหว่างวันที่ 1 มกราคม 2559 ถึง 5 พฤษภาคม 2566

Web Traffic Data from Google Analytics



Web Traffic Data from Google Analytics

ตัวอย่างชุดข้อมูล

Day Index	Users	Bounce Rate	Avg. Session Du	Number of Sess	Pages / Session	Pageviews	Sessions
1/1/16	1,777	47.83%	0:01:30	1.06	4.53	8,553	1,888
1/2/16	1,402	50.48%	0:01:49	1.04	6.42	9,359	1,458
1/3/16	1,333	52.40%	0:02:10	1.09	5.35	7,804	1,458
1/4/16	1,777	48.54%	0:03:57	1.08	9.91	18,981	1,916
1/5/16	1,736	47.70%	0:03:10	1.04	8.42	15,191	1,805
1/6/16	1,777	40.75%	0:01:30	1.06	5.78	10,844	1,875
1/7/16	1,805	46.35%	0:02:38	1.05	7.91	14,941	1,888
1/8/16	1,625	39.81%	0:02:44	1.05	7.49	12,788	1,708

การจัดเตรียมข้อมูล

- เปลี่ยนรูปแบบข้อมูล
 - เปลี่ยน data type ให้เหมาะสมกับงาน เนื่องจากเมื่อเก็บข้อมูลมาได้ ข้อมูลส่วนใหญ่จะอยู่ในรูปแบบ String
- ตั้งชื่อคอลัมน์
 - ตั้งชื่อคอลัมน์ให้เป็นไปตามข้อกำหนดของแบบจำลองนั้นๆ เช่น Prophet จะใช้คอลัมน์วันที่เป็น 'ds' และคอลัมน์ที่จะทำนายเป็น 'y'
- ลบข้อมูลที่ไม่จำเป็นออก

การจัดเตรียมข้อมูล

- สร้างตัวแปรใหม่ (Feature Engineering)
 - เป็นกระบวนการเพื่อสร้างคุณลักษณะ (features) หรือตัวแปรที่เหมาะสมและมีประโยชน์ในการสร้างแบบจำลอง (models) หรือวิเคราะห์ข้อมูล ซึ่งคุณลักษณะเหล่านี้จะถูกนำไปใช้ในการประมวลผลและการตัดสินใจในการแยกแยะข้อมูล มีเป้าหมายเพื่อเพิ่มประสิทธิภาพของแบบจำลอง ลดความซับซ้อน และเพิ่มความเข้าใจในข้อมูล

การจัดเตรียมข้อมูล

Feature Engineering

- **is_weekend**
 - สร้างตัวแปรขึ้นมาใหม่โดยอ้างอิงจากวันที่ เพื่อแสดงว่าวันที่เหล่านั้นเป็นวันเสาร์-อาทิตย์หรือไม่ โดยที่จะแสดงผลเป็น 1 ถ้าไม่ใช่วันเสาร์-อาทิตย์ จะแสดงผลเป็น 0
- **week_no**
 - สร้างตัวแปรขึ้นมาใหม่โดยอ้างอิงจากวันที่ เพื่อแสดงว่าเป็นสัปดาห์ที่เท่าไรในแต่ละปี ซึ่งจะมีค่าอยู่ที่ระหว่าง 1-52

การจัดเตรียมข้อมูล

Feature Engineering

- **avg_session_duration**
 - รวบรวมข้อมูลระยะเวลาเฉลี่ยของการเข้าชมเว็บแต่ละครั้งเพิ่มเติม จาก Google Analytics ซึ่งจะมีหน่วยเป็นวินาที
- **e-commerce holidays**
 - สร้างตัวแปรขึ้นมาใหม่จากการค้นหาข้อมูลของวันที่ของวันเทศกาลสำหรับธุรกิจ e-commerce เพื่อช่วยเพิ่มรายละเอียดและการเรียนรู้ให้แบบจำลองในการพยากรณ์ได้ ที่เพิ่มเติมจากวันหยุดที่มีอยู่ในฐานข้อมูลเดิมของ Prophet

e-commerce holidays	วันที่
St. Patrick's Day	17 มีนาคม ของทุกปี
Easter	ในเดือนเมษายน วันที่ไม่แน่นอน
Halloween	31 ตุลาคม ของทุกปี
Singles' Day	11 พฤศจิกายน ของทุกปี
Black Friday	วันศุกร์ที่สี่ของเดือนพฤศจิกายน
Cyber Monday	วันจันทร์แรกที่ถัดจาก Black Friday
Super Saturday	วันเสาร์ก่อนที่จะถึงเทศกาลคริสต์มาส
Christmas Eve	24 ธันวาคม ของทุกปี
New Year's Eve	31 ธันวาคม ของทุกปี

สร้างเป็น DataFrame ไว้
ใช้ในแบบจำลอง โดยการ
ใช้ข้อมูลวันที่ดังกล่าว
ตั้งแต่ปี 2016-2023

การจัดเตรียมข้อมูล

ตัวอย่างชุดข้อมูลหลัง Preprocessing

	index	ds	y	avg_session_duration	week_no	is_weekend
2678	2678	2023-05-02	3202	185.98	18	0
2679	2679	2023-05-03	2808	176.39	18	0
2680	2680	2023-05-04	2842	194.42	18	0
2681	2681	2023-05-05	2598	160.21	18	0
2682	2682	2023-05-06	1981	134.34	18	1

การสร้างแบบจำลอง

- สร้างแบบจำลองพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วัน ด้วย ARIMA
- ใช้ค่า $p, d, q = 2, 1, 3$ ซึ่งเป็นค่าที่ได้มาจากการใช้ Auto ARIMA

Forecasted values for the next 14 days:

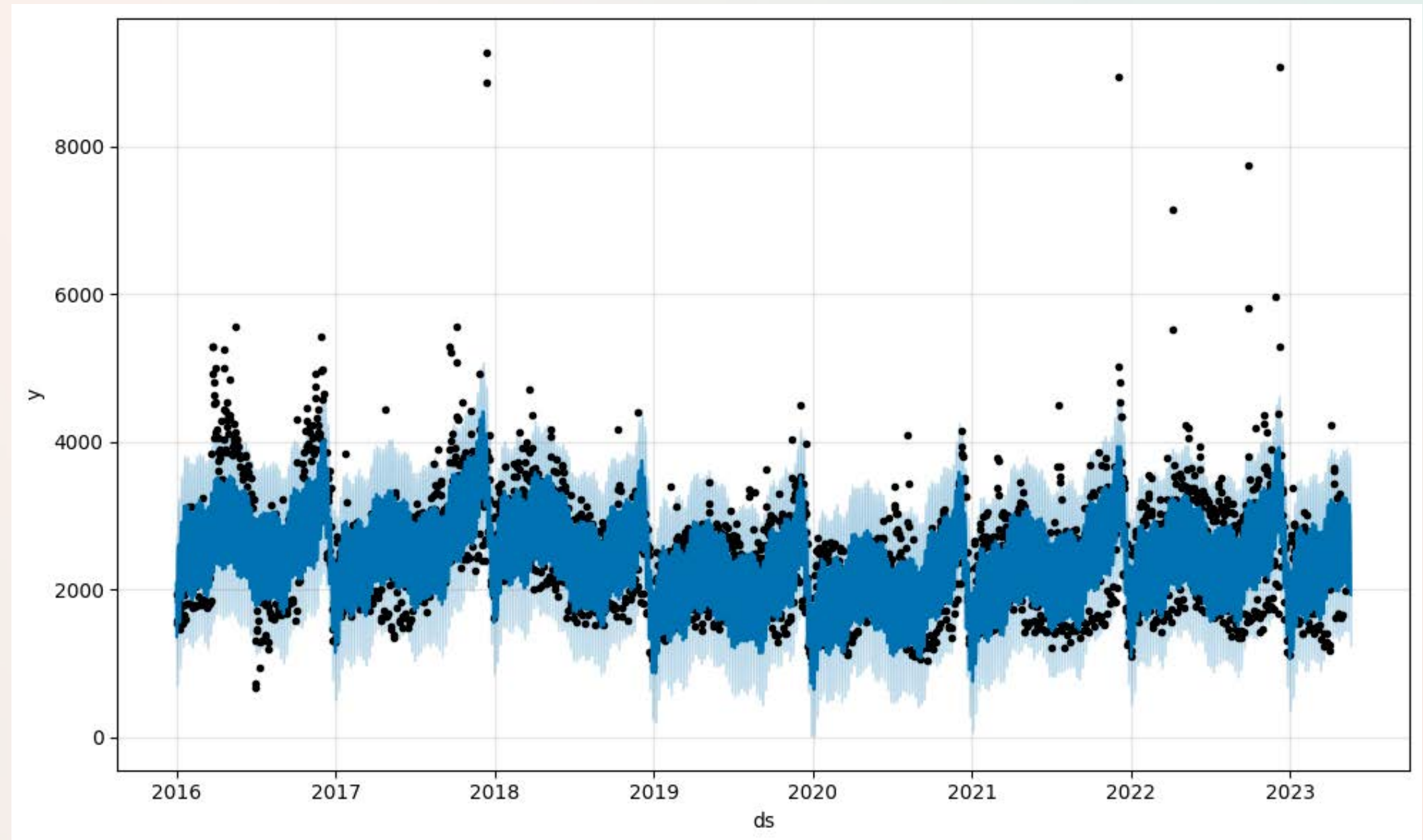
```
[2044.16762216 2577.46543803 3133.74068802 3294.1160154 2937.94220578  
2333.56161818 1936.13834057 2044.8703286 2577.73989516 3133.3803531  
3293.39239756 2937.40028642 2333.60940681 1936.7396857 ]
```


การสร้างแบบจำลอง

- สร้างแบบจำลองพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วัน ด้วย Prophet
- ปรับแต่งพารามิเตอร์ `seasonality_mode='multiplicative'` และ `make_future_dataframe(period=14, freq='D')`

ผลการพยากรณ์ของแบบจำลอง Prophet (Univariate)

	ds	yhat
2683	2023-05-07	2187.902042
2684	2023-05-08	3144.271320
2685	2023-05-09	3280.900947
2686	2023-05-10	3275.113125
2687	2023-05-11	3160.117909
2688	2023-05-12	2804.700363
2689	2023-05-13	2055.355926
2690	2023-05-14	2116.113016
2691	2023-05-15	3060.227099
2692	2023-05-16	3186.795175
2693	2023-05-17	3173.240219
2694	2023-05-18	3052.963201
2695	2023-05-19	2694.903159
2696	2023-05-20	1945.633185



การสร้างแบบจำลอง

- สร้างแบบจำลองพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วัน ด้วย Prophet แบบ Multivariate
- ปรับแต่งพารามิเตอร์ seasonality_mode='multiplicative' และ make_future_dataframe(period=14, freq='D')
- เพิ่มเติมในส่วนของการสร้าง holidays และ e-commerce holidays เข้ามาตามพารามิเตอร์ได้แก่ make_holidays_df(country='US')
- ใช้ฟังก์ชัน add_regressor เพื่อที่จะนำตัวแปรที่สร้างขึ้นมาใหม่มาช่วยในการพยากรณ์

ตัวอย่างชุดข้อมูลก่อนที่ จะทำการพยากรณ์

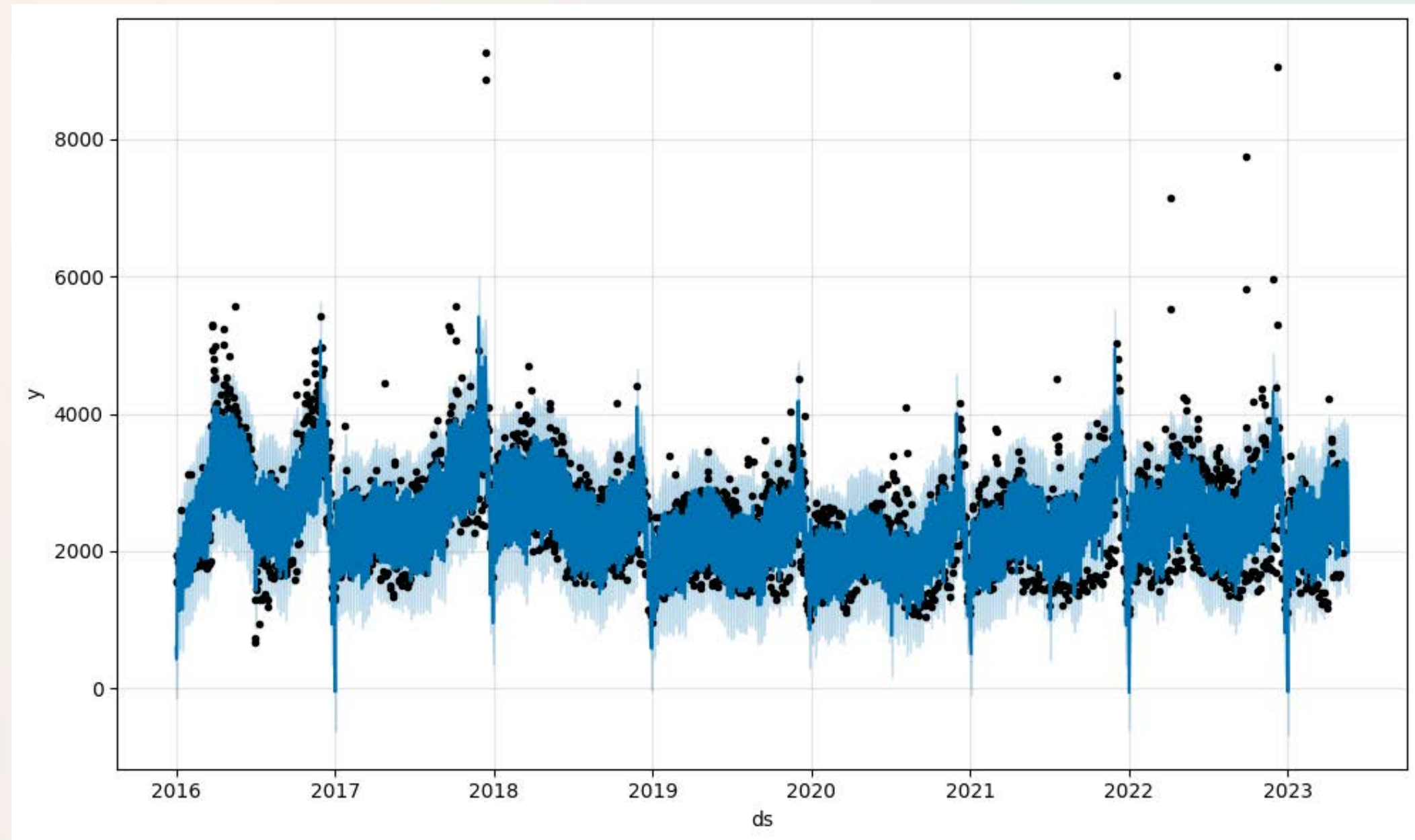
	ds	y	avg_session_duration	is_weekend	week_no
2677	2023-05-01	2468.0	178.480000	0	18
2678	2023-05-02	3202.0	185.980000	0	18
2679	2023-05-03	2808.0	176.390000	0	18
2680	2023-05-04	2842.0	194.420000	0	18
2681	2023-05-05	2598.0	160.210000	0	18
2682	2023-05-06	1981.0	134.340000	1	18
2683	2023-05-07	NaN	170.272065	1	18
2684	2023-05-08	NaN	170.272065	0	19
2685	2023-05-09	NaN	170.272065	0	19
2686	2023-05-10	NaN	170.272065	0	19
2687	2023-05-11	NaN	170.272065	0	19
2688	2023-05-12	NaN	170.272065	0	19
2689	2023-05-13	NaN	170.272065	1	19
2690	2023-05-14	NaN	170.272065	1	19
2691	2023-05-15	NaN	170.272065	0	20
2692	2023-05-16	NaN	170.272065	0	20
2693	2023-05-17	NaN	170.272065	0	20
2694	2023-05-18	NaN	170.272065	0	20
2695	2023-05-19	NaN	170.272065	0	20
2696	2023-05-20	NaN	170.272065	1	20

รายการ Holidays และ e-commerce holidays ที่ใช้แบบจำลอง

0	New Year's Day
1	Martin Luther King Jr. Day
2	Washington's Birthday
3	St. Patrick's Day
4	Easter
5	Memorial Day
6	Independence Day
7	Labor Day
8	Columbus Day
9	Halloween
10	Single Day
11	Veterans Day
12	Thanksgiving
13	Black Friday
14	Cyber Monday
15	Super Saturday
16	Christmas Eve
17	Christmas Day
18	Christmas Day (Observed)
19	New year eve
20	New Year's Day (Observed)
21	Veterans Day (Observed)
22	Independence Day (Observed)
23	Juneteenth National Independence Day (Observed)
24	Juneteenth National Independence Day

ผลการพยากรณ์ของแบบจำลอง Prophet (Multivariate)

	ds	yhat
2683	2023-05-07	2157.288511
2684	2023-05-08	3218.784425
2685	2023-05-09	3334.157834
2686	2023-05-10	3331.209260
2687	2023-05-11	3238.225519
2688	2023-05-12	2890.155165
2689	2023-05-13	2041.105600
2690	2023-05-14	2139.438604
2691	2023-05-15	3188.219417
2692	2023-05-16	3292.888774
2693	2023-05-17	3280.676759
2694	2023-05-18	3180.088680
2695	2023-05-19	2826.438551
2696	2023-05-20	1974.222634



การนำไปใช้งาน

- เป็นการแจ้งเตือนล่วงหน้า 14 วัน ผ่านแอปพลิเคชันไลน์ (Line)
- เมื่อค่า \hat{y} สูงกว่าเกณฑ์ที่ตั้งไว้จะทำการส่งการแจ้งเตือนโดยอัตโนมัติ
- ใช้วิธีการหา Outlier จาก Interquartile range (IQR) เพื่อนำมาใช้ในการกำหนดเกณฑ์ ซึ่งจะเลือกใช้ค่า Upper bound

$$\text{Upper bound} = Q3 + (\text{IQR} * 1.5)$$

LINE API Ecosystem



LINE Login



LINE Notify



Messaging API



LIFF



LINE Pay



LINE Things
LINE Beacon



Social API



LINE TV

การแจ้งเตือนผ่านแอปพลิเคชันไลน์ (Line)

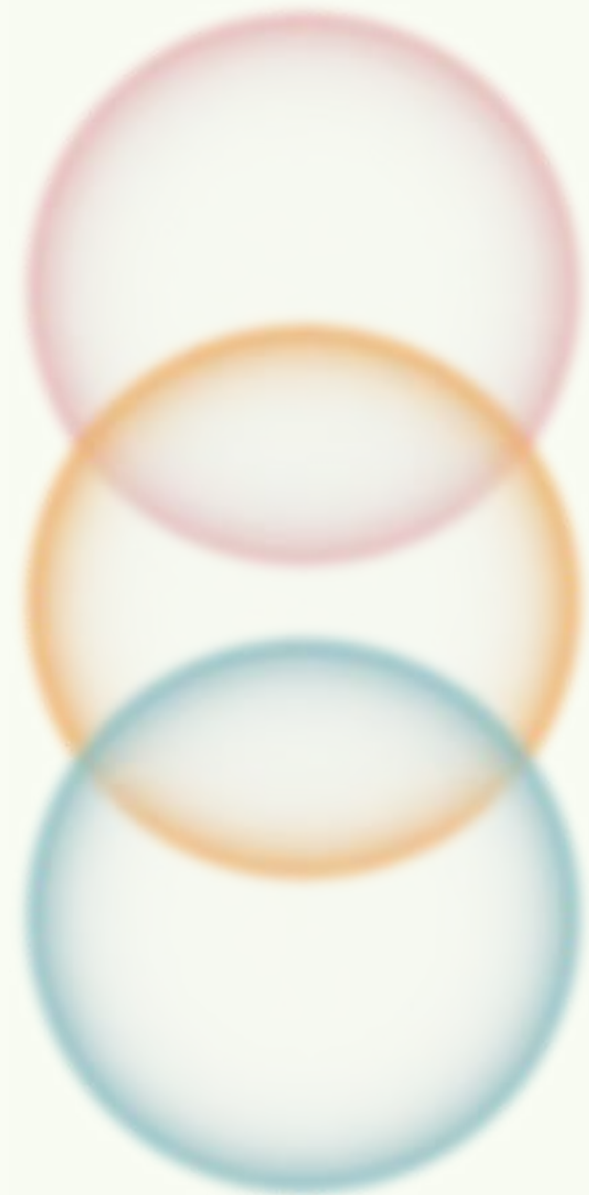


LINE Notify

SabaPing888: Local server on 2023-05-20 tends to be insufficient due to the predicted peak load reaching 1974.22 🔥 🔥 🔥 . Please be prepared



12:49 AM



ผลการวิจัย

ARIMA

Avg. RMSE: 617.336
Avg. MAE: 542.852
Avg. MAPE: 0.238

Horizon	RMSE	MAE	MAPE
1 day	63.168	63.168	0.032
2 days	575.544	575.149	0.257
3 days	809.479	657.036	0.29
4 days	816.197	719.306	0.309
5 days	758.835	666.633	0.259
6 days	432.06	419.788	0.158
7 days	262.949	208.036	0.09
8 days	524.485	433.95	0.204
9 days	790.917	701.154	0.335
10 days	967.215	838.617	0.397
11 days	973.001	842.239	0.388
12 days	826.681	737.679	0.317
13 days	560.77	506.388	0.202
14 days	281.397	230.79	0.095

Prophet

Univariate

Avg. RMSE: 434.898

Avg. MAE: 357.948

Avg. MAPE: 0.173

Horizon	RMSE	MAE	MAPE
2 days	322.849	267.576	0.151
3 days	500.68	412.833	0.238
4 days	471.871	385.048	0.19
5 days	384.408	332.434	0.152
6 days	411.181	356.904	0.18
7 days	444.612	403.361	0.231
8 days	378.911	328.384	0.185
9 days	541.692	395.526	0.166
10 days	564.653	445.163	0.19
11 days	472.972	387.102	0.175
12 days	337.286	280.647	0.126
13 days	410.776	334.005	0.135
14 days	411.785	324.345	0.13

Prophet

Multivariate

Avg. RMSE: 338.927

Avg. MAE: 275.686

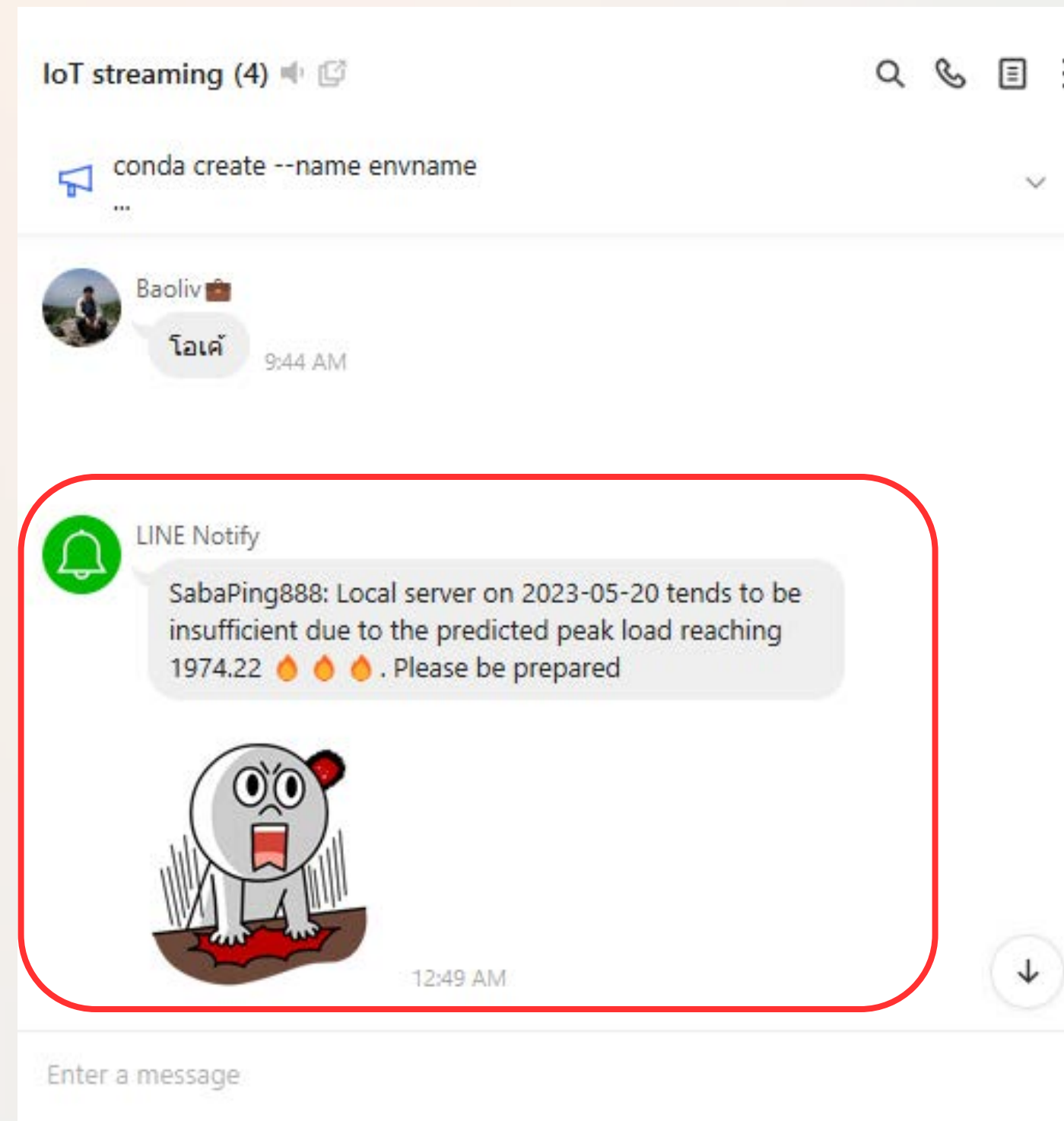
Avg. MAPE: 0.136

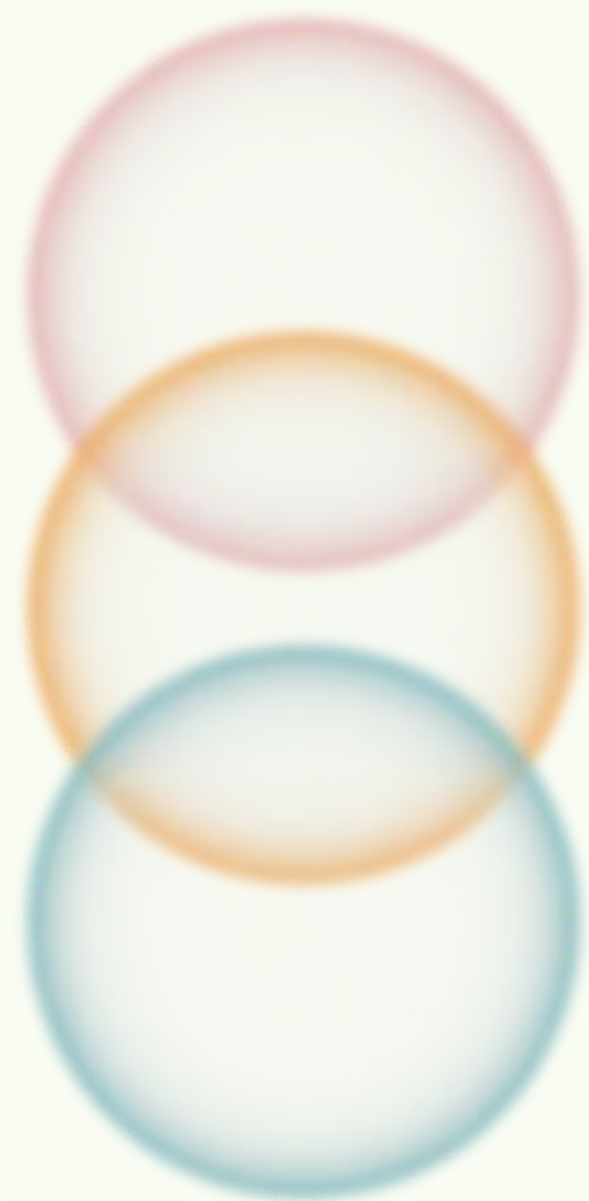
Horizon	RMSE	MAE	MAPE
2 days	421.397	331.615	0.18
3 days	371.928	295.82	0.172
4 days	290.346	235.363	0.128
5 days	326.201	291.792	0.147
6 days	448.867	365.931	0.191
7 days	360.739	300.145	0.164
8 days	277.244	245.521	0.134
9 days	396.295	314.753	0.142
10 days	355.742	279.775	0.13
11 days	303.217	233.536	0.101
12 days	270.91	219.458	0.092
13 days	304.04	245.407	0.1
14 days	279.129	224.805	0.09

ผลการวิจัย

จากตารางผลการทดสอบของโมเดล Prophet (Multivariate) มีประสิทธิภาพในการพยากรณ์ปริมาณการเข้าใช้เว็บไซต์ใน 14 วันล่วงหน้า ได้ดีที่สุด โดยที่ค่า RMSE เท่ากับ 279.129 ค่า MAE เท่ากับ 224.805 และ MAPE เท่ากับ 9% และภาพรวมของค่าเฉลี่ย Error น้อยกว่าแบบจำลองอื่นๆ จึงเลือกใช้งาน Prophet (Multivariate) เป็นแบบจำลองในการสร้างระบบแจ้งเตือน

ตัวอย่างการส่งข้อความแจ้งเตือนในกลุ่มสนทนา





สรุปผลและข้อเสนอแนะ

สรุปผลการศึกษา

ผลการทดลองพบว่าในการประเมินประสิทธิภาพของแบบจำลองที่ใช้ในการวิเคราะห์อนุกรมเวลาสำหรับการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ พบว่า Prophet แบบ Multivariate มีประสิทธิภาพที่ดีกว่า ARIMA และ Prophet แบบ Univariate การใช้วิธี ARIMA ในการวิเคราะห์อนุกรมเวลาเป็นวิธีที่ได้รับความนิยมในอดีต แต่ในการทดลองนี้พบว่า Prophet แบบ Multivariate มีประสิทธิภาพที่ดีกว่า ARIMA โดยสามารถให้ผลการพยากรณ์ที่แม่นยำขึ้นและเหมาะสมกับข้อมูลการเข้าชมเว็บไซต์มากขึ้น และในการทดลองนี้ ยังพบว่า Prophet แบบ Multivariate ให้ผลการพยากรณ์ที่ดีกว่า Prophet แบบ Univariate ซึ่งหมายความว่า การนำเข้าข้อมูลเพิ่มเติมเช่นตัวแปรอื่นๆ ที่สอดคล้องกับปริมาณการเข้าชมเว็บไซต์ สามารถช่วยเพิ่มประสิทธิภาพในการทำนายได้

ข้อเสนอแนะ

- ชุดข้อมูลที่นำมาใช้งานเป็นข้อมูลของเว็บไซต์ Google Merchandise Store ซึ่งทาง Google เปิดให้บริการสำหรับขายของที่ระลึก รวมทั้งยังใช้งานเป็นเว็บไซต์ตัวอย่างสำหรับการศึกษาวิธีใช้งาน Google Analytics ข้อมูลจึงอาจมีความคลาดเคลื่อน
- พิจารณาหาตัวแปรหรือปัจจัยอื่นๆที่ส่งผลต่อการพยากรณ์ นอกเหนือจากวันหยุดและ e-commerce holidays ได้แก่ กิจกรรมส่งเสริมการขายและการตลาดของแพลตฟอร์ม เทศกาลอื่นๆที่อาจจะส่งผลต่อการเข้าชมเว็บไซต์ เช่น Pride Month เป็นต้น
- สามารถพัฒนาประสิทธิภาพของแบบจำลองได้ด้วยการทดลองปรับค่าพารามิเตอร์ การสร้างตัวแปรใหม่ๆ หรือทดลองกับแบบจำลองประเภทอื่น เพื่อให้ได้ผลการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ที่แม่นยำมากยิ่งขึ้น

ขอขอบคุณครับ

Q&A