

การพยากรณ์ปริมาณการเข้าชมเว็บไซต์สำหรับธุรกิจอีคอมเมิร์ซ
โดยใช้เทคนิคการเรียนรู้ของเครื่องและสร้างระบบแจ้งเตือนโดยอัตโนมัติ

FORECASTING WEB TRAFFIC FOR E-COMMERCE WEBSITE
USING MACHINE LEARNING AND AUTOMATED PUSH NOTIFICATIONS

สิทธิ ใจวัฒนา¹
เอกสิทธิ์ พชรวงศ์ศักดิ์²

บทคัดย่อ

งานวิจัยนี้นำเสนอการพยากรณ์ปริมาณการเข้าชมเว็บไซต์สำหรับธุรกิจ e-commerce โดยใช้เทคนิคการเรียนรู้ของเครื่องในการพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting) ซึ่งจะช่วยคาดการณ์แนวโน้มปริมาณการเข้าชมเว็บไซต์ในอนาคต เมื่อปริมาณการเข้าชมเว็บไซต์เกินกว่าเกณฑ์ที่กำหนดไว้ ระบบจะแจ้งเตือนผ่านแอปพลิเคชันให้ผู้ดูแลระบบรับทราบเพื่อทำการปรับปรุงเซิร์ฟเวอร์ทันที นี้จะช่วยให้สามารถจัดการทรัพยากรพยากรณ์และเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ นอกจากนี้การพยากรณ์ปริมาณการเข้าชมเว็บไซต์อย่างแม่นยำ ยังสามารถใช้ในการวางแผนการตลาดและโปรแกรมส่งเสริมการขาย ช่วยให้ธุรกิจสามารถจัดสรรทรัพยากรได้อย่างเหมาะสมและเพิ่มประสบการณ์ใช้งานที่ดีสำหรับผู้ใช้งาน

คำสำคัญ : ปริมาณการเข้าชมเว็บไซต์, การเรียนรู้ของเครื่อง, อนุกรมเวลา, อีคอมเมิร์ซ

ABSTRACT

This research presents a framework for time series forecasting of web traffic on an e-commerce website, with a focus on optimizing server resource allocation and enhancing performance. Leveraging time series forecasting techniques, the study predicts future web traffic patterns, enabling businesses to efficiently manage server infrastructure and allocate resources effectively. The framework incorporates automated push notifications, alerting the system administrator when the forecasted web traffic exceeds a predefined

¹ นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่

² ที่ปรึกษาสารนิพนธ์

threshold. This proactive approach allows for timely server upscaling to ensure optimal website performance. The accurate forecasting of web traffic also facilitates strategic planning for marketing campaigns, enabling businesses to allocate resources efficiently. By leveraging time series forecasting, businesses can improve operational efficiency, maintain competitiveness, and deliver a seamless user experience on their e-commerce websites.

Keywords: Web Traffic, Machine learning, Time Series, E-Commerce

บทนำ

ณ ปัจจุบันในยุคที่เต็มไปด้วยเทคโนโลยีและความสะดวกสบายในการเชื่อมต่ออินเทอร์เน็ต การทำธุรกิจออนไลน์หรือ e-commerce เป็นส่วนสำคัญและเติบโตอย่างรวดเร็วในเศรษฐกิจโลก การพยากรณ์ปริมาณการเข้าชมเว็บไซต์เป็นสิ่งสำคัญที่ไม่สามารถละเว้นได้ในการที่จะประสบความสำเร็จในการดำเนินธุรกิจ e-commerce

การพยากรณ์ปริมาณการเข้าชมเว็บไซต์ช่วยให้ธุรกิจสามารถวางแผนและดำเนินการในอนาคตอย่างมีประสิทธิภาพ การพยากรณ์ที่แม่นยำช่วยให้ธุรกิจสามารถจัดสรรทรัพยากรและการตลาดให้เหมาะสม และช่วยให้การตัดสินใจทางธุรกิจเป็นไปตามความต้องการและการเติบโตของตลาด e-commerce อย่างมีประสิทธิภาพ

การพยากรณ์ปริมาณการเข้าชมเว็บไซต์ในอนาคตเป็นที่สนใจสำหรับธุรกิจ e-commerce เนื่องจากสามารถนำข้อมูลไปใช้ในการพัฒนาและปรับปรุงผลิตภัณฑ์และบริการ การพยากรณ์ที่แม่นยำช่วยให้ธุรกิจสามารถตอบสนองความต้องการของลูกค้าได้อย่างเหมาะสมและสร้างความพึงพอใจ นอกจากนี้ยังช่วยให้ธุรกิจสามารถตระหนักถึงโอกาสในการเติบโตของธุรกิจ e-commerce อย่างมีประสิทธิภาพ

ในด้านของการทำธุรกิจออนไลน์ การพยากรณ์ปริมาณการเข้าชมเว็บไซต์มาวิเคราะห์ข้อมูล จะสามารถทำให้เข้าใจพฤติกรรมของผู้ใช้งานเว็บไซต์มากยิ่งขึ้น

งานวิจัยนี้จึงนำเสนอการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ในอนาคตด้วยวิธีอนุกรมเวลาและส่งข้อความแจ้งเตือนผ่านทางแอปพลิเคชัน

วัตถุประสงค์ของงานวิจัย

1. ศึกษาความสามารถและประสิทธิภาพของ Prophet ในการทำนายการเข้าชมเว็บไซต์ในภาค e-commerce การศึกษานี้จะวิเคราะห์และทดสอบความแม่นยำและประสิทธิภาพของ Prophet ในการทำนายปริมาณการเข้าชมเว็บไซต์ e-commerce โดยเปรียบเทียบผลลัพธ์ที่ได้กับโมเดลทางสถิติอื่นๆ
2. เพื่อให้แนวทางและคำแนะนำที่เกี่ยวข้องสำหรับธุรกิจ e-commerce ผลลัพธ์จากการวิจัยนี้จะมี ความสำคัญในการสร้างความเข้าใจและแนะนำแนวทางการปฏิบัติที่มีประสิทธิภาพสำหรับธุรกิจที่ต้องการ ปรับปรุงกระบวนการทำนายปริมาณการเข้าชมเว็บไซต์ ซึ่งสามารถใช้เพื่อเพิ่มประสิทธิภาพในการวางแผนกล ยุทธ์การตลาด เป็นการจัดสรรทรัพยากรให้พอดีต่อการใช้งาน และช่วยให้ธุรกิจสามารถตอบสนองความ ต้องการของลูกค้าเพื่อที่จะเติบโตในตลาด e-commerce ได้อย่างมีประสิทธิภาพ

ขอบเขตงานวิจัย

1. ข้อมูลการเข้าใช้งานเว็บไซต์ Google Merchandise Store ตั้งแต่วันที่ 1 มกราคม 2559 ถึงวันที่ 5 พฤษภาคม 2566
2. เตรียมข้อมูลให้พร้อมสำหรับการนำมาวิเคราะห์และสร้างโมเดลสำหรับการพยากรณ์ปริมาณการเข้าชม เว็บไซต์
3. สร้างโมเดลการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ 14 วันล่วงหน้า
4. สร้างระบบแจ้งเตือนล่วงหน้า 14 วัน เพื่อให้ผู้เกี่ยวข้องได้มีเวลามากพอที่จะดำเนินการ

ทฤษฎี และผลงานที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำการพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting) มา ประยุกต์ใช้ในการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้าสำหรับธุรกิจ e-commerce โดยมีผลลัพธ์เป็นการ สร้างระบบแจ้งเตือนผ่านทางแอปพลิเคชัน จึงจำเป็นต้องศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

1. ความหมายและประเภทของ Web Traffic

Web traffic หมายถึงจำนวนข้อมูลที่ถูกรับและส่งผ่านเครือข่ายอินเทอร์เน็ตของเว็บไซต์ใน ระยะเวลาที่กำหนด โดยมักใช้เพื่อวัดปริมาณการเข้าชมของเว็บไซต์ในแต่ละวัน จำนวน web traffic สามารถสื่อ ถึงความนิยมและความสำเร็จของเว็บไซต์ได้ในบางครั้ง ยิ่งมีจำนวน web traffic มากขึ้น มักแปลว่ามีผู้เข้าชม

หรือผู้ใช้งานมากขึ้นเช่นกัน ซึ่ง Web Traffic มี 6 ประเภท ดังนี้ Organic Traffic, Referral Traffic, Direct Traffic, Social Media Traffic, Paid Traffic และ Campaign Traffic

2. ความหมายและประเภทของการพยากรณ์

การพยากรณ์เป็นกระบวนการที่ใช้ข้อมูลที่มีอยู่ในอดีตหรือปัจจุบันเพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคต ซึ่งสามารถนำไปใช้ในการวิเคราะห์เพื่อให้เห็นแนวโน้มของการเปลี่ยนแปลงในสิ่งที่เราต้องการพยากรณ์ โดยใช้การวิเคราะห์และการสร้างโมเดลทางสถิติและคณิตศาสตร์ เพื่อประมาณค่าหรือเหตุการณ์ที่อาจเกิดขึ้นในอนาคต โดยสามารถนำไปใช้ในการวางแผนสำหรับการดำเนินงาน อีกทั้งยังช่วยในการประกอบการตัดสินใจเพื่อให้บรรลุวัตถุประสงค์ของการดำเนินการ การพยากรณ์แบ่งเป็น 2 ประเภทคือ 1) การพยากรณ์ตามระยะเวลา ได้แก่ การพยากรณ์ระยะสั้น การพยากรณ์ระยะกลาง และการพยากรณ์ระยะยาว 2) การพยากรณ์ตามคุณลักษณะ ได้แก่ วิธีการเชิงคุณภาพ และวิธีการเชิงปริมาณ

3. พยากรณ์แบบอนุกรมเวลา (Time Series Forecasting)

ในการศึกษานี้จะขอทำการยกตัวอย่างจากการเก็บข้อมูลปริมาณการเข้าชมเว็บไซต์และกิจกรรมอื่นๆ ภายในกรอบระยะเวลาหนึ่งสม่ำเสมอ โดยมีช่วงห่างในการเก็บข้อมูลเท่ากัน องค์ประกอบของอนุกรมเวลา (Time Series Components) เทคนิคการพยากรณ์แบบอนุกรมเวลา มีสมมติฐานว่าปริมาณการเข้าชมเว็บไซต์ มีการเปลี่ยนแปลงตามระยะเวลาที่ผ่านมา ดังนั้นการคำนวณด้วยสมการทางคณิตศาสตร์ ตัวแปรที่มีผลต่อปริมาณการเข้าชมเว็บไซต์คือเวลา (Time) โดยทั่วไปปริมาณการเข้าชมเว็บไซต์ เกิดจากปัจจัย 4 อย่าง ดังนี้ แนวโน้ม (Trend), ฤดูกาล (Seasonal), วัฏจักร (Cycle) และความผิดปกติ (Irregular)

แนวโน้ม (Trend) หมายถึง ปริมาณการเข้าชมเว็บไซต์มีลักษณะเพิ่มขึ้นหรือลดลงอย่างต่อเนื่องไป โดยส่วนมากอิทธิพลของแนวโน้มมีสาเหตุมาจากเหตุการณ์ต่างๆ การเปลี่ยนแปลงกลุ่มเป้าหมายของเว็บไซต์ การโฆษณา เทคโนโลยีที่มีผลต่อการเข้าชม รวมทั้งกิจกรรมทางการตลาดของเว็บไซต์

ฤดูกาล (Seasonal) หมายถึง ปริมาณการเข้าชมเว็บไซต์ที่เปลี่ยนแปลงมีลักษณะเป็นรูปแบบซ้ำๆ กันจากผลกระทบของฤดูกาล โดยสามารถสังเกตเห็นรูปแบบ Pattern ที่ชัดเจนและเกิดตามรูปแบบนั้นๆ ซ้ำในช่วงเวลาถัดมา โดยหน่วยของระยะเวลาเป็นได้ทั้งรายปี รายเดือน รายวัน หรือรายชั่วโมง ส่วนมากลักษณะปริมาณการเข้าชมเว็บไซต์แบบฤดูกาลมีอิทธิพลมาจากเทศกาลหรือวันหยุดต่างๆ

วัฏจักร (Cycle) หมายถึง ปริมาณการเข้าชมเว็บไซต์ที่มีลักษณะลดลงหรือเพิ่มขึ้น โดยไม่เกี่ยวข้องกับฤดูกาล สามารถมีระยะเวลายาวเท่าหรือยาวกว่าฤดูกาล เช่น การลดลงหรือเพิ่มขึ้นของผู้เข้าชมเว็บไซต์ในระยะยาวเท่ากับหรือยาวกว่าหนึ่งปี

ความผิดปกติ (Irregular) หมายถึง ความเป็นธรรมชาติของข้อมูลที่มีความผิดปกติหรือสิ่งที่ไม่สามารถคาดเดาได้ล่วงหน้า อาจเกิดจากปัจจัยที่ไม่สามารถควบคุมได้หรือเหตุการณ์ที่ไม่คาดคิด เช่น ภาวะฉุกเฉิน

สภาวะภัยพิบัติ หรือสภาวะเหตุการณ์ไม่สงบ เป็นต้น ซึ่งสามารถทำให้เกิดการเปลี่ยนแปลงของปริมาณการเข้าชมเว็บไซต์

4. Autoregressive Integrated Moving Average (ARIMA)

ARIMA (Autoregressive Integrated Moving Average) เป็น โมเดลเชิงสถิติที่ใช้เพื่อการพยากรณ์ และวิเคราะห์อนุกรมเวลา (Time Series Analysis) โดยเฉพาะ โดยมีความสามารถในการจัดการกับองค์ประกอบของแนวโน้ม (Trend) และสภาวะซ้ำซ้อน (Seasonality) ซึ่งเป็นความสามารถที่สำคัญในการพยากรณ์และการวิเคราะห์แนวโน้มของข้อมูลตามเวลาในอนาคต ARIMA ประกอบด้วยสามส่วนหลัก คือ AR (Autoregressive), I (Integrated), และ MA (Moving Average)

5. Prophet

Prophet เป็นโอเพนซอร์สสำหรับการพยากรณ์ข้อมูลแบบอนุกรมเวลา สามารถใช้งานได้ทั้งบน R และ Python ซึ่งถูกพัฒนาโดยกลุ่มนักวิทยาศาสตร์ข้อมูลภายในองค์กร Facebook (ที่รู้จักในปัจจุบันว่า Meta) ได้นำมาใช้ในงานพยากรณ์ภายใน Facebook ในหลายโครงการ และได้ทำการปล่อยให้ใช้งานกันแบบสาธารณะในปี 2017 อัลกอริทึมของ Prophet จะทำงานได้ดีกับข้อมูลที่มีหลายฤดูกาล มีปริมาณข้อมูลมาก และอธิบายคุณลักษณะของฤดูกาลของข้อมูลได้

Sean J. Taylor และ Ben Letham คือกลุ่มผู้พัฒนา Prophet ซึ่งมีความต้องการให้ Prophet เป็นโมเดลที่ใช้งานได้ง่าย โดยที่ไม่จำเป็นต้องมีความรู้เกี่ยวกับรายละเอียดของโมเดลพื้นฐานมาก อธิบายผลได้ง่าย โดยการแบ่งการทำงานออกเป็นส่วนใหญ่ๆ ซึ่งมีคุณลักษณะเด่นๆของ Prophet มีดังนี้

มีความยืดหยุ่น คือสามารถจัดการกับงานพยากรณ์อนุกรมเวลาได้หลายรูปแบบ เช่น ข้อมูลที่ไม่มี ความสม่ำเสมอหรือข้อมูลที่ขาดหาย จัดการกับ outlier ได้ดี เป็นต้น

การจำแนกแนวโน้ม คือสามารถตรวจจับและจัดการแนวโน้มในข้อมูลได้หลายรูปแบบ ไม่ว่าจะเป็น Linear หรือ Non-Linear รวมถึงการเปลี่ยนแปลงของแนวโน้ม

การจำแนกฤดูกาล คือการที่ Prophet สามารถจัดการรูปแบบฤดูกาลได้หลายช่วงเวลา เช่น รายวัน รายสัปดาห์ รายเดือน หรือรายปี โดยที่จะนำรูปแบบดังกล่าวไปใช้ในการพยากรณ์

การวิเคราะห์ผลกระทบจากวันหยุด คือช่วยให้ผู้ใช้ระบุเหตุการณ์ วันหยุด เทศกาล ที่กำหนดเองและ คาดว่าจะส่งผลกระทบต่อข้อมูล เพื่อให้โมเดลได้รับรู้ถึงความแตกต่างระหว่างฤดูกาลหรือวัฏจักรที่เกิดจากวันหยุดหรือเหตุการณ์พิเศษ

การแสดงผลและการวิเคราะห์ข้อมูล คือความสามารถในการแสดงผลการพยากรณ์ ออกมาในรูปแบบกราฟต่างๆ เช่น แนวโน้มของผลพยากรณ์รายวัน รายสัปดาห์ รายเดือน และรายปี, แสดงความสำคัญของวันหยุดต่างๆที่ส่งผลต่อการพยากรณ์ เป็นต้น

Prophet มีพื้นฐานมาจากการวิเคราะห์ จาก 3 ส่วนประกอบหลัก ซึ่งอ้างอิงมาจาก Additive model คือ แนวโน้มของพฤติกรรมของข้อมูล (Trend), พฤติกรรมของข้อมูลที่ไม่ได้เป็นเชิงเส้นจะถูกอธิบายด้วยฤดูกาล (Seasonality) ในช่วงระยะเวลารายปี รายเดือน รายวัน, และส่วนสุดท้ายคือ ผลกระทบที่เกิดจากวันหยุด (Effect of holidays) ซึ่งจะช่วยให้โมเดลเรียนรู้ ช่วยทำให้มีความแม่นยำสูงขึ้น

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

โดยที่ $g(t)$ คือ แนวโน้ม (Trend)
 $s(t)$ คือ ฤดูกาล (Seasonality)
 $h(t)$ คือ ผลกระทบจากวันหยุด (Effect of holidays)
 et คือ ความผิดปกติหรือความไม่แน่นอน (Error)

Prophet นำเสนอโมเดลสำหรับการพยากรณ์ข้อมูลแบบอนุกรมเวลาที่ใช้ saturating growth model และ piecewise linear model สำหรับการวิเคราะห์แนวโน้ม (Trend), ใช้อัลกอริทึมฟูเรียร์เป็นฐานในการวิเคราะห์ฤดูกาล (Seasonality) เพื่อใช้ในการสร้างโมเดลที่ยืดหยุ่นในการตรวจจับและวิเคราะห์ฤดูกาลของชุดข้อมูล และในส่วนของการกระทบจากวันหยุด (Effect of holidays) ที่มีผลต่อข้อมูล นอกเหนือจากวันหยุดที่เกิดขึ้นซ้ำในช่วงเวลาเดียวกันของปี ยังมีเหตุการณ์จำนวนมากและวันหยุดที่มีวันที่ไม่แน่นอนในแต่ละปี ที่ส่งผลกระทบต่อพฤติกรรมของข้อมูล นอกจากนี้วันหยุดยังแตกต่างกันตามแต่ละประเทศซึ่งอ้างอิงจากปฏิทินจันทรคติ Prophet ให้ความสำคัญกับวันหยุดสากลและวันหยุดของแต่ละประเทศ และสมมติให้เหตุการณ์ วันหยุด เป็นอิสระต่อกัน และนำมาใช้งานคล้ายกับการวิเคราะห์ฤดูกาล

6. งานวิจัยที่เกี่ยวข้อง

Tejas Shelatkar, Stephen Tondale, Swaraj Yadav, and Sheetal Ahir, (2020) ได้ นำ เ ส น อ ก ร พยากรณ์ข้อมูลอนุกรมเวลาด้วยการ ใช้ Discrete Wavelet Transform (DWT) แบ่งข้อมูลที่ได้มาจาก Wikipedia Pageview API ออกเป็น 2 ส่วน ในขั้นตอนการจัดเตรียมข้อมูล ทำให้ได้ส่วนที่เป็น Detail และ Approximate เพื่อที่จะแบ่งให้ ARIMA และ LSTM RNN ช่วยกันพยากรณ์ และจะนำผลการพยากรณ์กลับมารวมกันด้วยวิธี Invert Discrete Wavelet Transform (iDWT) ทำให้ได้ผลการพยากรณ์ที่แม่นยำมากขึ้น

D. Sikka and C. N. S. Vinoth Kumar (2023) ได้นำเสนอการพยากรณ์ข้อมูลอนุกรมเวลาด้วยการใช้ ensemble model ที่ประกอบไปด้วยโมเดลหลายแบบได้แก่ Decision Tree, Multiple Linear Regression และ Support Vector Machine ในการพยากรณ์ข้อมูล pageview และ visitors ในการหาจำนวน unique visitors

N. Petluri and E. Al-Masri (2018) ได้นำเสนอการพยากรณ์อนุกรมเวลาด้วยวิธีการสร้างตัวแปรใหม่ (Feature Engineering) จากโมเดลเดิมที่มีคือ RNN seq2seq โดยที่ผู้วิจัยได้ตั้งสมมติฐานขึ้นมาว่าการสร้างตัวแปรใหม่ขึ้นมา จะช่วยส่งเสริมการเรียนรู้ให้กับโมเดล ซึ่งจะทำให้ประสิทธิภาพของโมเดลที่มีอยู่นั้นสูงขึ้นได้

ระเบียบวิธีวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำการพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting) มาประยุกต์ใช้ในการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้าสำหรับธุรกิจ e-commerce โดยมีผลลัพธ์เป็นการสร้างระบบแจ้งเตือนผ่านทางแอปพลิเคชัน โดยมีขั้นตอนการดำเนินการดังนี้

1. การรวบรวมข้อมูล

งานวิจัยนี้ได้ใช้ข้อมูล Web Traffic จาก Google Merchandise Store ตั้งแต่วันที่ 1 มกราคม 2559 ถึง 5 พฤษภาคม 2566 ซึ่งถูกรวบรวมโดย Google Analytics ซึ่งเป็นเครื่องมือวิเคราะห์และรายงานข้อมูลให้บริการโดย Google

2. การจัดเตรียมข้อมูล

2.1 การเปลี่ยนรูปแบบข้อมูล

เปลี่ยน data type ให้เหมาะสมกับงาน เนื่องจากข้อมูลที่มีส่วนใหญ่ จะอยู่ในรูปแบบ String

2.2 การตั้งชื่อคอลัมน์

ตั้งชื่อคอลัมน์ให้เป็นไปตามข้อกำหนดของแบบจำลองนั้นๆ เช่น Prophet จะใช้คอลัมน์วันที่เป็น “ds” และคอลัมน์ที่จะทำนายเป็น “y”

2.3 การสร้างตัวแปรใหม่

เป็นกระบวนการเพื่อสร้างคุณลักษณะ (features) หรือตัวแปรที่เหมาะสมและมีประโยชน์ในการสร้างแบบจำลอง (models) หรือวิเคราะห์ข้อมูล ซึ่งคุณลักษณะเหล่านี้จะถูกนำไปใช้ในการประมวลผลและการตัดสินใจในการแยกแยะข้อมูล มีเป้าหมายเพื่อเพิ่มประสิทธิภาพของแบบจำลอง ลดความซับซ้อน และเพิ่มความเข้าใจในข้อมูล

3. การสร้างแบบจำลอง

3.1 สร้างแบบจำลองพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วันด้วย ARIMA ใช้ค่า $p, d, q = 2, 1, 3$ ซึ่งเป็นค่าที่ได้มาจากการใช้ Auto ARIMA

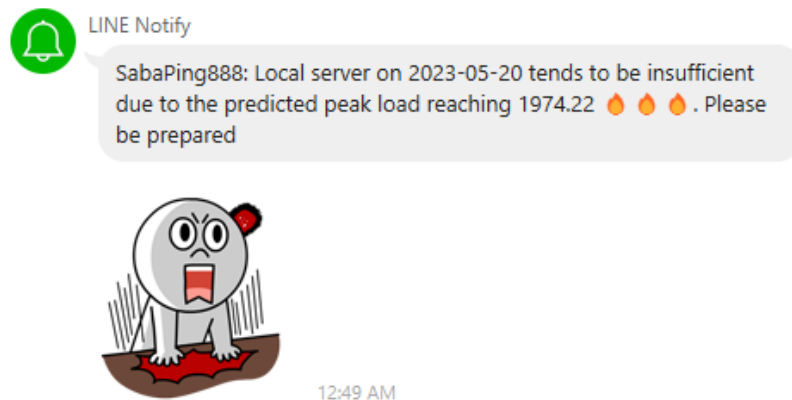
3.2 สร้างแบบจำลองพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วันด้วย Prophet

3.3 สร้างแบบจำลองพยากรณ์ปริมาณการเข้าชมเว็บไซต์ล่วงหน้า 14 วันด้วย Prophet แบบ Multivariate เพิ่มเติมในส่วนของการสร้าง holidays และ e-commerce holidays เข้ามาในการประมวลผล และใช้ฟังก์ชัน add regressor เพื่อที่จะนำตัวแปรที่สร้างขึ้นมาใหม่มาช่วยในการพยากรณ์

3.4 วัดประสิทธิภาพของผลลัพธ์ที่ได้ว่ามีความน่าเชื่อถือมากน้อยเพียงใด โดยทำการวัดผลจากค่าความคลาดเคลื่อน (Error) ต่างๆ ได้แก่ Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) และ Root Mean Square Error (RMSE)

3.5 กำหนดเกณฑ์ที่จะใช้ในการนำไปใช้ในการแจ้งเตือน งานวิจัยนี้เลือกใช้วิธีการหา Outlier จาก Interquartile range (IQR) เพื่อที่นำมาใช้ในการกำหนดเกณฑ์ ซึ่งจะเลือกใช้ค่า upper bound คือการนำค่าใน คอวไทล์ที่ 3 มาบวกกับ 1.5 เท่าของ IQR

3.6 นำโมเดลไปใช้งานในการแจ้งเตือนผู้เกี่ยวข้อง เมื่อค่าที่พยากรณ์ได้มีค่าสูงกว่าเกณฑ์ที่กำหนดไว้โดยเป็นการส่งข้อความผ่านแอปพลิเคชันไลน์ (Line) ด้วยการใช้ API ของ Line Notify



ภาพที่ 1 ตัวอย่างผลลัพธ์การแจ้งเตือนผ่านทางแอปพลิเคชันไลน์ (Line)

4. เครื่องมือที่ใช้ในงานวิจัย

งานวิจัยนี้ ใช้ภาษาไพทอนตั้งแต่ขั้นตอนการเตรียมข้อมูล (Data processing) ไปจนถึง ขั้นตอนการสร้างแบบจำลอง และใช้ Line Notify API ในการส่งข้อความแจ้งเตือนอัตโนมัติ

ผลการศึกษา

งานวิจัยนี้ได้ทำการทดลองสร้างแบบจำลอง ARIMA, Prophet แบบ Univariate และ Prophet แบบ Multivariate ได้ผลดังตารางที่ 1 ถึงตารางที่ 3

ตารางที่ 1 ตารางแสดงผลการทดสอบประสิทธิภาพของโมเดล ARIMA

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	0.290
4 days	816.197	719.306	0.309
5 days	758.835	666.633	0.259
6 days	432.060	419.788	0.158
7 days	262.949	208.036	0.090
8 days	524.485	433.950	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	560.770	506.388	0.202
14 days	281.397	230.790	0.095

ตารางที่ 2 ตารางแสดงผลการทดสอบประสิทธิภาพของโมเดล Prophet แบบ Univariate

Horizon	RMSE	MAE	MAPE
2 days	322.849	267.576	0.151
3 days	500.680	412.833	0.238
4 days	471.871	385.048	0.190
5 days	384.408	332.434	0.152
6 days	411.181	356.904	0.180
7 days	444.612	403.361	0.231

ตารางที่ 2 (ต่อ)

Horizon	RMSE	MAE	MAPE
8 days	378.911	328.384	0.185
9 days	541.692	395.526	0.166
10 days	564.653	445.163	0.190
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	0.130

ตารางที่ 3 ตารางแสดงผลการทดสอบประสิทธิภาพของโมเดล Prophet แบบ Multivariate

Horizon	RMSE	MAE	MAPE
2 days	421.397	331.615	0.180
3 days	371.928	295.820	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	0.130
11 days	303.217	233.536	0.101
12 days	270.910	219.458	0.092
13 days	304.040	245.407	0.100
14 days	279.129	224.805	0.090

จากตารางผลการทดสอบของโมเดล Prophet (Multivariate) มีประสิทธิภาพในการพยากรณ์ปริมาณการเข้าใช้เว็บไซต์ใน 14 วันล่วงหน้า ได้ดีที่สุด โดยที่ค่า RMSE เท่ากับ 279.129 ค่า MAE เท่ากับ 224.805 และ MAPE เท่ากับ 9% และภาพรวมของค่า Error ต่างๆมีความนิ่งกว่าโมเดลอื่นๆ

สรุปและอภิปรายผลการวิจัย

จากผลการทดลองพบว่าในการประเมินประสิทธิภาพของโมเดลที่ใช้ในการวิเคราะห์อนุกรมเวลาสำหรับการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ พบว่า Prophet แบบ Multivariate มีประสิทธิภาพที่ดีกว่า ARIMA และ Prophet แบบ Univariate ยังพบว่ากรณีที่ Prophet แบบ Multivariate ให้ผลการพยากรณ์ที่ดีกว่า Prophet แบบ Univariate หมายความว่า การนำเข้าข้อมูลเพิ่มเติม เช่น ตัวแปรอื่นๆ ที่สอดคล้องกับปริมาณการเข้าชมเว็บไซต์ สามารถช่วยเพิ่มประสิทธิภาพในการทำนายได้

นำผลลัพธ์ที่ได้จากโมเดล ไปพัฒนาเป็นระบบการแจ้งเตือนอัตโนมัติ ในกรณีที่ผลการพยากรณ์ล่วงหน้า 14 วัน สูงกว่าเกณฑ์ที่กำหนด เพื่อเป็นตัวช่วยในการตัดสินใจให้แก่ผู้ดูแลระบบ

ข้อเสนอแนะ

1. ชุดข้อมูลที่นำมาใช้งานเป็นข้อมูลของเว็บไซต์ Google Merchandise Store ซึ่งทาง Google เปิดให้บริการซื้อขายของที่ระลึก รวมทั้งยังใช้งานเป็นเว็บไซต์ตัวอย่างในการศึกษาวิธีใช้งาน Google Analytics ข้อมูลจึงอาจมีความคลาดเคลื่อนบ้างในบางวัน
2. พิจารณาหาตัวแปรหรือปัจจัยอื่นๆ ที่ส่งผลกระทบต่อผลการพยากรณ์ นอกเหนือจากวันหยุดและ e-commerce holidays เช่น เทศกาลอื่นๆ ที่อาจจะส่งผลกระทบต่อปริมาณการเข้าชมเว็บไซต์ เช่น Pride Month เป็นต้น
3. สามารถพัฒนาประสิทธิภาพของโมเดลได้ด้วยการทดลองปรับค่าพารามิเตอร์ การสร้างตัวแปรใหม่ๆ หรือทดลองกับโมเดลประเภทอื่น เพื่อให้ได้ผลการพยากรณ์ปริมาณการเข้าชมเว็บไซต์ที่มีความแม่นยำสูงขึ้น

บรรณานุกรม

- Petluri, N. & Al-Masri, E. (2018), Web Traffic Prediction of Wikipedia Pages, *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 5427-5429.
doi: 10.1109/BigData.2018.8622207.
- Rafferty, G. (2023). *Forecasting Time Series Data with Prophet - Build, improve, and optimize time series forecasting model using Meta's advanced forecasting tool*. Packt Publishing.
- Shukla, S. (2023). *Various Techniques to Detect and Isolate Time Series Components Using Python*. <https://www.analyticsvidhya.com/blog/2023/02/various-techniques-to-detect-and-isolate-time-series-components-using-python/>
- Shelatkar, T., Tondale, S., Yadav, S., & Ahir, S. (2020). Web traffic time series forecasting using ARIMA and LSTM RNN. *ITM Web of Conferences* (Vol. 32, p. 03017). EDP Sciences.

- Sikka, D. & C. N. S. Vinoth Kumar (2023), Website Traffic Time Series Forecasting Using Regression Machine Learning, *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, 2023, pp. 246-250.
doi: 10.1109/CSNT57126.2023.10134631.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), pp. 37-45, Sep. 2017. doi: <https://doi.org/10.7287/peerj.preprints.3190v2>.
- Wangpratham, N. (2563, 20 เมษายน). การพยากรณ์ข้อมูลอนุกรมเวลาด้วยเทคนิค ARIMA ด้วย Python. <https://medium.com/qunt-i-love-u/การพยากรณ์ข้อมูลอนุกรมเวลาด้วยเทคนิค-arima-ด้วย-python-44809eb8e990>
- Zunic, E., Korjenic, K., Delalic, S., & Subara, Z. (2021). Comparison Analysis of Facebook's Prophet, Amazon's DeepAR+ and CNN-QR Algorithms for Successful Real-World Sales Forecasting. *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 13, no. 2, Apr. 2021. doi: 10.5121/ijcsit.2021.13205.