A FORECASTING MODEL FOR HALF-HOURLY INBOUND CALL VOLUME IN A CALL CENTER SERVICE

Kritchaya PRAPHARUT[1]

1 Faculty of Engineering, Chulalongkorn University, Thailand; kritchaya.ph@gmail.com

ARTICLE HISTORY

Received: 19 April 2024 **Revised:** 3 May 2024 **Published:** 17 May 2024

ABSTRACT

The objective of this research is to develop the model for interval call forecasting for a call center service company. Call center service company is the main contact point for customer service channel to serve many requirements including enquiries, suggestions, and any cases solving. The company needs to plan the call center agent workforce to match with expected incoming call. Sometimes, the plan needs to change according to unmet between actual incoming calls and forecasting. The current forecasting method can potentially be improved since it has low accuracy and can't re-forecast on a daily basis. This research proposes the model to better forecast interval incoming call with the objective to forecast call after 10 am of the day to support workforce management during the day. The model concept is to do clustering then forecast both the patterns and number of calls. In conclusion, the model has MAPE at 20.8% which is a lot lower than the company's current method with MAPE at 52.7%.

Keywords: Call Center, Interval Call Forecasting, Machine Learning

CITATION INFORMATION: Prapharut, K. (2024). A Forecasting Model for Half-Hourly Inbound Call Volume in a Call Center Service. *Procedia of Multidisciplinary Research, 2*(5), 30

การพยากรณ์จำนวนสายโทรศัพท์เข้าของศูนย์บริการข้อมูลทางโทรศัพท์แบบรายครึ่งชั่วโมง

กฤตชญา ประภารัตน์[1]

1 คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย; kritchaya.ph@gmail.com

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาโมเดลที่เหมาะสมสำหรับการพยากรณ์จำนวนสายโทรศัพท์เข้าของศูนย์บริการข้อมูลทางโทรศัพท์แบบรายครึ่งชั่วโมง โดยศูนย์บริการข้อมูลทางโทรศัพท์ หรือ Call Center มีบทบาทเป็นศูนย์รวมสายโทรเข้าและโทรออกของธุรกิจ ซึ่งเป็นช่องทางสำคัญในการตอบสนองความต้องการของลูกค้า ไม่ว่าจะเป็นการสอบถามข้อมูล การขอคำแนะนำ หรือแก้ปัญหาต่างๆ ศูนย์บริการข้อมูลทางโทรศัพท์จึงมีการจัดวางแผนกำลังคนรับสาย เพื่อให้สอดคล้องกับปริมาณสายโทรศัพท์ที่คาดว่าจะเข้ามา แต่ในบางครั้งการวางแผนจัดกำลังคนรับสายอาจต้องมีการปรับระหว่างวัน เนื่องจากจำนวนสายโทรเข้าอาจมีจำนวนมากกว่าหรือน้อยกว่าที่คาดการณ์ไว้ ซึ่งวิธีการเดิมที่บริษัทใช้ในการคำนวน อาจมีความคลาดเคลื่อน และไม่สามารถปรับตัวเลขได้ภายในระยะเวลาอันสั้น งานวิจัยนี้จึงนำเสนอการพยากรณ์จำนวนสายโทรศัพท์เข้าของศูนย์บริการข้อมูลทางโทรศัพท์แบบรายครึ่งชั่วโมง โดยมีวัตถุประสงค์เพื่อพยากรณ์ปริมาณสายการโทรเข้าช่วงหลัง 10 น. เพื่อช่วยให้ฝ่ายวางแผนกำลังคนเห็นแนวโน้มปริมาณสายที่คาดว่าจะเข้ามา และตัดสินใจปรับแผนการจัดกำลังคนได้อย่างทันท่วงที โดยโมเดลจะจัดกลุ่มและพยากรณ์รูปแบบการกระจายตัวของปริมาณสายโทรเข้า และพยากรณ์จำนวนสายที่คาดว่าจะเข้ามา โดยผลการทดลองพบว่า โมเดลที่พัฒนาขึ้นมามีความแม่นยำมากกว่าวิธีการคำนวนเดิมของบริษัท โดย MAPE ของโมเดลที่พัฒนาขึ้นอยู่ที่ 20.8% ซึ่งดีกว่าวิธีการของบริษัทซึ่ง MAPE อยู่ที่ 52.7%

คำสำคัญ: ศูนย์บริการข้อมูลทางโทรศัพท์, การพยากรณ์จำนวนสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง, การเรียนรู้ของเครื่องกล

ข้อมูลการอ้างอิง: กฤตชญา ประภารัตน์. (2567). การพยากรณ์จำนวนสายโทรศัพท์เข้าของศูนย์บริการข้อมูลทางโทรศัพท์แบบรายครึ่งชั่วโมง. *Procedia of Multidisciplinary Research, 2*(5), 30

บทนำ

ศูนย์บริการข้อมูลทางโทรศัพท์ (Call Center) เป็นช่องทางในการตอบสนองความต้องการของลูกค้า ไม่ว่าจะเป็นเรื่องการสอบถามข้อมูล การขอคำแนะนำ หรือการแก้ปัญหาต่างๆ ซึ่งเป็นส่วนหนึ่งที่สะท้อนภาพลักษณ์ของธุรกิจถึงเรื่องความใส่ใจ และการบริการ มีบทบาทเป็นศูนย์รวมการโทรเข้าและโทรออกของธุรกิจ โดยทั้งสองรูปแบบแตกต่างกันตรงที่ การโทรเข้า จำเป็นต้องวางกำลังคนรับสายให้เพียงพอ โดยต้องคำนึงถึงต้นทุนและจำนวนพนักงานรับสายที่มีโดยปริมาณสายที่โทรเข้าจะมีลักษณะแตกต่างกันไปตามช่วงวันและเวลา ดังนั้นการพยากรณ์ปริมาณสายการโทรเข้าจึงเป็นส่วนสำคัญในการดำเนินงานของศูนย์บริการข้อมูลทางโทรศัพท์ ในปัจจุบันมีงานวิจัยที่ทำโมเดลมาเพื่อรองรับการพยากรณ์ปริมาณสายการโทรเข้า เช่น งานของ Zhang et al. (2021) ที่ได้นำเสนอการใช้โมเดล The Deep Neural Network Consists of a Long-Short Term Memory (LSTM) และ Autoregressive Integrated Moving Average (ARIMA) มาเปรียบเทียบประสิทธิภาพกัน และงานของ Kumwilaisak (2022) ที่ใช้ โมเดล The Deep Neural Network Consists of a Long-Short Term Memory และ Network and a Deep Neural Network (DNN) มาใช้ในการจัดกำลังคนรับสาย (Kumwilaisak et al., 2022) แม้ว่าการพยากรณ์ปริมาณจำนวนสายโทรเข้าที่แม่นยำจะเป็นเรื่องสำคัญเพื่อวางแผนจัดกำลังคนล่วงหน้า แต่ธุรกิจศูนย์บริการข้อมูลทางโทรศัพท์มักมีผลกระทบจากปัจจัยภายนอกหรือปัจจัยบางประการที่ไม่ทราบล่วงหน้า เช่น สภาพอากาศ วันสำคัญ หรือประกาศต่างๆ ส่งผลให้อาจเกิดกรณีที่มีจำนวนสายโทรเข้ามามากกว่าหรือน้อยกว่าที่คาดการณ์ไว้ เพื่อช่วยให้ฝ่ายวางแผนจัดกำลังคนรับสายเห็นแนวโน้มจำนวนสายที่คาดว่าจะเข้ามาภายในวันนั้นและสามารถทำการปรับแผนจัดกำลังคนรับสายได้อย่างทันท่วงที งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาโมเดลสำหรับพยากรณ์ปริมาณสายการโทรเข้าช่วงหลัง 10 น. ซึ่งเป็นเวลามาตรฐานที่ศูนย์บริการข้อมูลทางโทรศัพท์ใช้ตัดสินใจปรับตัวเลขจำนวนคนรับสายภายในวันนั้น เพื่อช่วยให้ศูนย์บริการข้อมูลทางโทรศัพท์สามารถปรับเปลี่ยนแผนการจัดกำลังคนรับสายได้อย่างเหมาะสม และช่วยลดต้นทุนที่เกิดขึ้น

การทบทวนวรรณกรรม

การศึกษางานวิจัยในครั้งนี้ ผู้วิจัยได้ทบทวนวรรณกรรมที่เกี่ยวข้องเพื่อนำมากำหนดแนวทางการวิจัย รวมถึงโมเดลที่จะเลือกไปศึกษา เพื่อพัฒนาโมเดลที่เหมาะสำหรับพยากรณ์ปริมาณสายการโทรเข้าเพื่อปรับกำลังคนรับสายในช่วงระหว่างวัน โดยใช้ข้อมูลช่วงก่อน 10 น. มาพยากรณ์จำนวนสายที่คาดว่าจะเข้ามาในช่วงหลัง 10 น. จากงานวิจัยของ Zhang (2021) และ Kumwilaisak (2022) ซึ่งได้กล่าวถึงโมเดล LSTM ผู้วิจัยจึงได้นำโมเดล LSTM มาศึกษา พบว่าโมเดล LSTM เป็นโมเดลที่ใช้กับข้อมูลแบบ Time Series จึงไม่เหมาะกับข้อมูลของงานวิจัยนี้ที่มีลักษณะ Cross Sectional ผู้วิจัยจึงมาศึกษาโมเดลที่ใช้พยากรณ์ปัญหาในลักษณะ Regression Problem แทน

สำหรับโมเดลที่ใช้พยากรณ์ปัญหาในลักษณะ Regression Problem นั้น จากงานวิจัยของ García-Gutiérrez et al. (2015) ได้มีการศึกษาเพื่อเปรียบเทียบโมเดลที่เหมาะสมในการพยากรณ์ความสูงของพื้นที่ป่า โดยเปรียบเทียบ 4 โมเดล ได้แก่ 1) Neural Networks 2) Support Vector Regression (SVR) 3) KNN และ 4) Random Forest Regression ซึ่งโมเดลที่มีประสิทธิภาพสูงที่สุดคือ SVR ในขณะที่ Bag (2020) และ Acharya et al. (2019) ได้ทำการวิจัยเพื่อเปรียบเทียบโมเดลในการพยากรณ์ค่าความน่าจะเป็นที่จะสอบเข้ามหาวิทยาลัยได้ งานวิจัยของ Bag (2020) ได้ทำการเปรียบเทียบทั้งหมด 3 โมเดล ได้แก่ 1) Linear Regression 2) Random Forest Regression 3) Decision Tree Regression และในขณะที่ Acharya et al. (2019) ได้เพิ่ม Support Vector Regression เข้ามาในการเปรียบเทียบ จากทั้ง 2 งานวิจัยพบว่า โมเดล Linear Regression เป็นโมเดลที่มีประสิทธิภาพสูงสุดในการพยากรณ์ค่าความน่าจะเป็นที่จะสอบเข้ามหาวิทยาลัยได้ ดังนั้น ผู้วิจัยจึงได้นำโมเดล Linear Regression มาทำการทดลอง พบว่า ค่า MAPE อยู่ที่ 41.1% ซึ่งเป็นผลมาจากการที่ข้อมูลมีลักษณะการกระจายตัวของปริมาณสายการโทรเข้าในแต่ละวันแตกต่างกัน การนำข้อมูลรายวันมาพยากรณ์รวมกันจึงทำให้เกิดความไม่แม่นยำ ดังนั้น ผู้วิจัยจึงนำโมเดลการแบ่งกลุ่ม (Clustering) มาใช้ร่วมกับงานวิจัยในครั้งนี้ เพื่อทำการแบ่งกลุ่มวันตามลักษณะการกระจายตัวของปริมาณสายการโทรเข้า และใช้โมเดล

สำหรับพยากรณ์กลุ่มของข้อมูลแบบหลายกลุ่ม (Multi-Class Classification) เพื่อพยากรณ์รูปแบบการกระจายตัวของปริมาณสายการโทรเข้าในแต่ละวัน ดังนั้น งานวิจัยชิ้นนี้จึงแบ่งโมเดลที่จะใช้ในการทดลองออกเป็น 3 ประเภท ได้แก่ โมเดลสำหรับทำ Clustering โมเดลสำหรับทำ Multi-Class Classification และโมเดลสำหรับทำ Regression จากศึกษางานวิจัยของ O'Neill et al. (2019) ได้ใช้ข้อมูลต่างๆ ของลูกค้าที่ติดต่อเข้ามา เช่น จำนวนสาย ระยะเวลาเฉลี่ยในการโทร และค่าเบี่ยงเบนมาตรฐานของระยะเวลาในการโทรจากจำนวนสายทั้งหมดของลูกค้าแต่ละคน ในการแบ่งกลุ่มลักษณะประเภทของลูกค้าโดยใช้โมเดล K-Mean ทำให้ได้กลุ่มลูกค้าที่มีพฤติกรรมแตกต่างกันอย่างชัดเจนตามลักษณะพฤติกรรมการติดต่อเข้ามา ดังนั้นโมเดล K-Mean จึงเป็นโมเดลหนึ่งที่ถูกเลือกนำมาทำ Clustering ในงานวิจัยชิ้นนี้

สำหรับการทำโมเดล Multi-Class Classification ได้มีผู้ทำการศึกษาเปรียบเทียบโมเดลทางด้าน Machine Learning ต่างๆ โดยงานวิจัยของ Çolakoğlu and Akkaya (2019) ได้ทำการเปรียบเทียบโมเดลในการจำแนกประเภทผู้ป่วย มีผู้ป่วยทั้งหมด 3 ประเภท และได้เปรียบเทียบทั้งหมด 9 โมเดล ได้แก่ 1) SVM 2) Naive Bayes (Gaussian) 3) KNN 4) Multilayer Perceptron (ANN) 5) Random Forest 6) CART (Classification and Regression Trees) 7) Logistic Regression 8) Gradient Boosting Machine 9) Extreme Gradient Boosting (XGBoost) จากงานวิจัยนี้ โมเดล Random Forest มีประสิทธิภาพสูงที่สุดในการจำแนกประเภทผู้ป่วย ในขณะที่ Khan et al. (2023) ได้ศึกษาเปรียบเทียบโมเดลที่ช่วยในการจำแนกประเภทถั่ว จากโมเดลทั้งหมด 8 โมเดล ได้แก่ 1) Logistic Regression 2) Naïve Bayes 3) KNN 4) Decision Tree 5) Random Forest 6) Extreme Gradient Boosting (XGBoost) 7) Support Vector Machine (SVM) 8) Multilayer Perception (MLP) จากการศึกษาของ Khan et al. (2023) พบว่า XGBoost เป็นโมเดลที่จำแนกประเภทได้ดีที่สุด ในขณะที่ Random Forest มีความแม่นยำในระดับที่สูงเช่นกัน ดังนั้นในการศึกษานี้ จึงได้นำโมเดล Random Forest และ XGBoost มาใช้ในการศึกษา รวมทั้งโมเดลอื่นๆ เพื่อนำมาเปรียบเทียบเพิ่มเติมอย่าง Logistics Regression

สำหรับการทำโมเดล Regression นั้น จากงานวิจัยของ García-Gutiérrez et al. (2015) และ Acharya et al. (2019) ที่กล่าวไปในช่วงต้น ผู้วิจัยจึงได้เลือกโมเดล SVR และ Linear Regression รวมถึงโมเดล Random Forest มาทำการทดลองในครั้งนี้ร่วมกัน

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวิธีการศึกษาตามกระบวนการวิเคราะห์ข้อมูลดังนี้

การทำความเข้าใจข้อมูล (Data Understanding) คือ ทำความเข้าใจข้อมูลตั้งต้นที่มาจากฐานข้อมูลของระบบให้บริการข้อมูลและทำรายการธุรกิจอัตโนมัติทางโทรศัพท์ (Interactive Voice Response: IVR)

การเตรียมข้อมูล (Data Preparation) คือ การดำเนินการคัดกรอง (Filter) เฉพาะข้อมูลที่เป็นสายโทรเข้า คัดเลือกตัวแปรที่มาจากระบบไอวีอาร์มาเป็นตัวแปรต้น 10 ตัวแปรที่สอดคล้องกับลักษณะธุรกิจและเหมาะที่จะนำมาเข้าโมเดล ได้แก่

1) จำนวนสายโทรเข้า
2) ระยะเวลาการรอสายของสายที่รอคิวที่นานที่สุดใน 1 interval
3) ระดับการให้บริการในการรับสาย (% Service Level)
4) อัตราส่วนของสายที่รับได้
5) อัตราส่วนของสายที่รับไม่ได้
6) ระยะเวลารวมเฉลี่ยในการให้บริการต่อ 1 สาย
7) ระยะเวลาพูดคุยเฉลี่ยในการให้บริการต่อ 1 สาย
8) ระยะเวลาถือสายรอเฉลี่ยระหว่างการให้บริการต่อ 1 สาย

9) ระยะเวลาบันทึกข้อมูลต่างๆ ของการให้บริการต่อ 1 สาย
10) ความเร็วเฉลี่ยที่ใช้ในการรับสาย
และจัดการข้อมูลให้มาอยู่ในรูปแบบที่เหมาะสมในการนำไปเข้าโมเดลสำหรับข้อมูลลักษณะตัดขวาง (Cross Sectional Data) หลังจากนั้นทำการแบ่งข้อมูลออกเป็น Train 60%, Validation 20% และ Test 20% เพื่อใช้ในการพัฒนาโมเดล
การแบ่งกลุ่มเพื่อสร้าง Label และรูปแบบการกระจายตัวของปริมาณสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง คือ การจัดวันที่มีลักษณะจำนวนสายโทรเข้าในช่วงแต่ละครึ่งชั่วโมงที่คล้ายกันมาอยู่กลุ่มเดียวกัน และนำผลการจัดกลุ่มไปเป็น Label และในส่วนของรูปแบบการกระจายตัวปริมาณสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง สร้างมาจากการหาค่าเฉลี่ยจำนวนสายโทรศัพท์ของข้อมูลแต่ละวันที่อยู่ในกลุ่มเดียวกัน เพื่อนำไปใช้เป็นรูปแบบสำหรับคูณกระจายค่าที่ได้จากการพยากรณ์จำนวนสายโทรศัพท์เข้าที่คาดว่าจะเข้ามาใน 1 วัน ให้อยู่ในรูปแบบรายครึ่งชั่วโมง
การสร้างโมเดล (Modeling) ในงานวิจัยนี้จะพัฒนาโมเดลทั้งหมด 2 ประเภท โดยในแต่ละประเภทจะพัฒนาโมเดลเพื่อใช้เปรียบเทียบหาโมเดลที่แม่นยำที่สุดในการนำไปพยากรณ์จำนวนสายโทรเข้า
1) โมเดลประเภทพยากรณ์ข้อมูลประเภทเชิงคุณภาพ (Classification) โดยจะสร้างโมเดลมาจากข้อมูล Label ที่ถูกจำแนกมาจาก Clustering และตัวแปรต้นต่างๆ ที่มีการเตรียมข้อมูลให้เป็นค่าของช่วงเวลารายครึ่งชั่วโมงในแต่ละวัน
2) โมเดลประเภทพยากรณ์ข้อมูลประเภทเชิงปริมาณ (Regression) โดยจะสร้างโมเดลจากข้อมูลจำนวนสายโทรศัพท์เข้าใน 1 วัน และตัวแปรต้นต่างๆ ที่มีการเตรียมข้อมูลให้คำนวนรวมเป็นค่าตัวแปรต้นใน 1 วัน ของแต่ละตัวแปรตาม
การวัดประสิทธิภาพของโมเดล (Model Evaluation) แบ่งตามประเภทโมเดล ได้แก่ โมเดลประเภท Classification จะใช้ค่า Accuracy และโมเดลประเภท Regression จะใช้ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error: MAPE)
โดยงานวิจัยนี้ ศึกษาภายใต้ขอบเขตของข้อมูลดังต่อไปนี้
1) ที่มาของข้อมูล: จากบริษัทที่ให้บริการทางด้านศูนย์บริการข้อมูลทางโทรศัพท์
2) ลักษณะธุรกิจ: ข้อมูลการติดต่อสอบถาม ร้องเรียน และหัวข้ออื่นๆ ที่เกี่ยวข้องกับข้อมูลทางด้านประกันสังคม
3) ลักษณะช่องทางการติดต่อ: เป็นข้อมูลเฉพาะสายโทรเข้า (Inbound) แต่ไม่รวมถึงจำนวนสายโทรออก (Outbound)
4) ระยะเวลาของข้อมูล: เป็นข้อมูลสายโทรเข้าในช่วง 1 มีนาคม 2021 ถึง 10 มิถุนายน 2022
5) ความถี่ในการวิเคราะห์: พยากรณ์จำนวนสายโทรเข้าที่จะเข้ามารวมทั้งหมดในช่วงเวลา 10 น. เป็นต้นไป (1 ช่วงเวลา = 30 นาที)

ผลการวิจัย

งานวิจัยนี้ผ่านการวิเคราะห์ในส่วนต่างๆ เพื่อพัฒนาโมเดล ซึ่งได้ผลการวิจัย ดังนี้
1) ขั้นตอนการแบ่งวันตามลักษณะการกระจายตัวของจำนวนสายโทรศัพท์เข้า (Clustering) เป็นขั้นตอนในการจัดกลุ่มวันที่มีลักษณะจำนวนสายโทรเข้าแบบรายครึ่งชั่วโมงที่ใกล้เคียงกันมาอยู่ด้วยกัน เพื่อใช้สร้างรูปแบบลักษณะการโทรเข้าของแต่ละกลุ่ม โดยใช้วิธีการจัดกลุ่มข้อมูลแบบเคมีน ผลจากการใช้วิธีดำเนินการจุดหักศอก (Elbow Method) ในการหาจำนวนกลุ่มที่เหมาะสม ได้ผลลัพธ์ว่าจุดที่เหมาะสมสำหรับแบ่งจำนวนกลุ่ม คือ 3 กลุ่ม เนื่องจากเป็นจุดที่กราฟมีลักษณะหักศอกมากที่สุด ผลลัพธ์จากการจัดกลุ่ม ได้ว่า จากข้อมูลทั้งหมดจำนวน 467 วัน ถูกจัดเป็นกลุ่มที่ 1 จำนวน 209 วัน จัดเป็นกลุ่มที่ 2 จำนวน 69 วัน และจัดเป็นกลุ่มที่ 3 จำนวน 189 วัน
2) ขั้นตอนการสร้างโมเดลพยากรณ์กลุ่มเพื่อหาลักษณะการกระจายตัวของจำนวนสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง มีการทดลองทั้งหมด 3 โมเดล ได้แก่ Logistic Regression, Random Forest Classification และ eXtreme Gradient Boosting (XGBoost) โดยการทดสอบประสิทธิภาพของโมเดลด้วยค่า Accuracy ซึ่งได้ผลลัพธ์ดังตารางที่ 1

ตารางที่ 1 แสดงผลการทดลองจากโมเดลพยากรณ์กลุ่มเพื่อหาลักษณะการกระจายตัวของจำนวนสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง

โมเดล	Accuracy
Logistic Regression	74%
Random Forest Classification	80%
eXtreme Gradient Boosting (XGBoost)	87%

3) ขั้นตอนการสร้างโมเดลที่ใช้พยากรณ์จำนวนสายโทรศัพท์เข้าใน 1 วัน มีการทดลองทั้งหมด 3 โมเดล ได้แก่ SGD Regression, Support Vector Machines Regression (SVR) และ Random Forest Regression โดยการทดสอบประสิทธิภาพของโมเดลด้วยค่า MAPE ซึ่งได้ผลลัพธ์ดังตารางที่ 2

ตารางที่ 2 แสดงผลการทดลองจากโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าใน 1 วัน

โมเดล	MAPE (Train Set)	MAPE (Validation Set)	MAPE (Test Set)
SGD Regression	14.39%	23.35%	26.90%
Support Vector Machines Regression (SVR)	7.00%	12.00%	9.00%
Random Forest Regression	2.60%	11.68%	7.88%

4) ขั้นตอนการพยากรณ์จำนวนสายโทรศัพท์เข้าที่คาดว่าจะเข้ามาในช่วงหลัง 10 น. เป็นต้นไป แบบรายครึ่งชั่วโมง จะเป็นการนำโมเดลพยากรณ์กลุ่มเพื่อหาลักษณะการกระจายตัวของจำนวนสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง โดยใช้ Hyperparameter ที่ได้ผลความแม่นยำดีที่สุดทั้ง 3 โมเดล มาทดลองใช้ร่วมกับโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าใน 1 วัน ทั้งหมด 3 โมเดล โดยใช้ Hyperparameters ที่ดีที่สุดเช่นกัน โดยเมื่อได้ผลจากโมเดลพยากรณ์กลุ่มว่าข้อมูลในวันถัดมาน่าจะมีรูปแบบการกระจายตัวของจำนวนสายโทรศัพท์เข้าตรงกับรูปแบบกลุ่มใด จากนั้นจึงนำผลพยากรณ์จำนวนสายโทรศัพท์ที่คาดว่าจะเข้ามาใน 1 วันนั้น มาคูณกระจายตามรูปแบบที่ได้พยากรณ์ไว้ในตอนแรก จึงจะได้ชุดข้อมูลการพยากรณ์จำนวนสายโทรศัพท์ที่คาดว่าจะเข้ามาแบบรายครึ่งชั่วโมงนั่นเอง ในการทดสอบประสิทธิภาพของโมเดลจะวัดด้วยค่า MAPE ซึ่งได้ผลลัพธ์ดังตารางที่ 3

ตารางที่ 3 แสดงผลการทดลองค่า MAPE จากโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าที่คาดว่าจะเข้ามาในช่วงหลัง 10 น.

Classification Model	Regression Model	MAPE
Logistic Regression	SGD Regression	40.80%
Logistic Regression	Support Vector Machines Regression	28.50%
Logistic Regression	Random Forest Regression	27.60%
Random Forest Classification	SGD Regression	32.70%
Random Forest Classification	Support Vector Machines Regression	21.40%
Random Forest Classification	Random Forest Regression	20.80%
eXtreme Gradient Boosting	SGD Regression	34.20%
eXtreme Gradient Boosting	Support Vector Machines Regression	21.70%
eXtreme Gradient Boosting	Random Forest Regression	21.02%

จากตารางพบว่า การทำงานของโมเดลพยากรณ์ลักษณะจำนวนสายโทรศัพท์เข้าแบบรายครึ่งชั่วโมง Random Forest Classification คู่กับโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าใน 1 วัน Random Forest Regression เป็นโมเดลที่ให้ค่า MAPE ต่ำที่สุด ซึ่งอยู่ที่ 20.80%

อภิปรายผลการวิจัย

งานวิจัยนี้ต้องการหาโมเดลที่เหมาะสมสำหรับพยากรณ์จำนวนสายโทรศัพท์เข้าที่คาดว่าจะเข้ามาในช่วงหลัง 10 น. เป็นต้นไป ของบริษัทศูนย์บริการข้อมูลทางโทรศัพท์ การศึกษาเริ่มต้นด้วยการแบ่งกลุ่มวันตามลักษณะการกระจายตัวของจำนวนสายโทรศัพท์เข้า ออกเป็นจำนวน 3 กลุ่ม จากนั้นจึงทำการหา Hyperparameter ที่เหมาะสมที่สุดเพื่อใช้สร้างโมเดลพยากรณ์ลักษณะการกระจายตัวจำนวนสายโทรศัพท์ และรวมถึงโมเดลพยากรณ์จำนวนสายโทรศัพท์ที่คาดว่าจะเข้ามาใน 1 วัน มาประกอบกันโดยการนำจำนวนสายโทรศัพท์ที่คาดว่าจะเข้ามาใน 1 วัน ไปคูณกระจายตามรูปแบบลักษณะการกระจายตัวที่ได้พยากรณ์ไว้

การทดลองได้มีการนำโมเดลมาทดลองร่วมกัน รวมทั้งสิ้น 12 โมเดล และได้นำมาเปรียบเทียบประสิทธิภาพของโมเดลด้วยค่า MAPE ได้ผลสรุปว่า การใช้โมเดลพยากรณ์กลุ่มเพื่อหาลักษณะจำนวนโทรศัพท์เข้าแบบรายครึ่งชั่วโมงเป็น Random Forest Classification คู่กับโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าใน 1 วัน เป็น Random Forest Regression เป็นโมเดลที่ให้ค่า MAPE ต่ำที่สุด ซึ่งอยู่ที่ 20.80% ซึ่งเมื่อเทียบกับการพยากรณ์ปริมาณสายการโทรเข้าโดยวิธีการคำนวนปัจจุบันของบริษัท จะมีค่า MAPE อยู่ที่ 52.70%

นอกจากนี้ ผู้วิจัยได้ทดลองทำการพยากรณ์ปริมาณสายการโทรเข้าโดยไม่ใช้เทคนิคการแบ่งกลุ่มวันเพื่อหารูปแบบการโทรเข้าแบบรายครึ่งชั่วโมง แต่ใช้เพียงโมเดลพยากรณ์ปริมาณสายการโทรเข้าแบบ Random Forest Regression พบว่า MAPE อยู่ที่ 42.23% จึงแสดงให้เห็นว่า การใช้เทคนิดแบ่งกลุ่มวันตามลักษณะจำนวนสายโทรศัพท์เข้าร่วมกับการทำโมเดลพยากรณ์กลุ่ม และโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าใน 1 วัน จะให้ความแม่นยำที่สูงกว่าและประสิทธิภาพที่ดีกว่าการทำเพียงโมเดลพยากรณ์จำนวนสายโทรศัพท์เข้าอย่างเดียว อีกทั้งยังให้ความแม่นยำที่ดีกว่าการใช้วิธีการคำนวนแบบปัจจุบันของบริษัทด้วยเช่นกัน

ข้อเสนอแนะในการวิจัยครั้งต่อไป

จากผลการวิจัย ผู้ที่สนใจสามารถนำโมเดล Random Forest Classification ประกอบกับโมเดล Random Forest Regression ซึ่งได้ผลลัพธ์ที่ดีที่สุดในการพยากรณ์จำนวนสายโทรเข้าของบริษัทศูนย์บริการข้อมูลทางโทรศัพท์ไปประยุกต์ใช้ให้กับบริษัทหรือหน่วยงานต่างๆ เช่น บริษัทศูนย์บริการข้อมูลทางโทรศัพท์ หรือหน่วยงานที่เกี่ยวข้องกับการให้บริการลูกค้าทางโทรศัพท์ภายในบริษัทหรือหน่วยงานต่างๆ โดยเฉพาะบริษัทหรือหน่วยงานที่อาจยังไม่มีการนำโมเดลทางสถิติหรือวิทยาการคอมพิวเตอร์มาประยุกต์ใช้ในการพยากรณ์จำนวนสายโทรเข้า นอกจากนี้ ผู้ที่สนใจสามารถนำผลลัพธ์จากงานวิจัยนี้ ไปประยุกต์ใช้ในบริบททางธุรกิจอื่นๆ ที่อาจมีลักษณะที่คล้ายกัน เช่น การพยากรณ์จำนวนการร้องเรียนในช่องทางต่างๆ นอกเหนือจากช่องทางโทรศัพท์

เอกสารอ้างอิง

Acharya, M. S., Armaan, A., & Antony, A. S. (2019). *A comparison of regression models for prediction of graduate admissions*. In 2019 international conference on computational intelligence in data science (ICCIDS) (pp. 1-5). IEEE.

Bag, A. (2020). A comparative study of regression algorithms for predicting graduate admission to a university. *Research Gate*.

Çolakoğlu, N., & Akkaya, B. (2019). *Comparison of multi-class classification algorithms on early diagnosis of heart diseases*. In y-BIS 2019 Conference Book: Recent Advances n Data Sc ence and Bus ness Analyt cs (p. 162).

Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. New York: Cambridge University Press.

García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., & Riquelme, J. C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing, 167*, 24-31.

Fan, J., & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Khan, M. S., Nath, T. D., Hossain, M. M., Mukherjee, A., Hasnath, H. B., Meem, T. M., & Khan, U. (2023). Comparison of multiclass classification techniques using dry bean dataset. *International Journal of Cognitive Computing in Engineering, 4*, 6-20.

Kumwilaisak, W., Phikulngoen, S., Piriyataravet, J., Thatphithakkul, N., & Hansakunbuntheung, C. (2022). Adaptive Call Center Workforce Management with Deep Neural Network and Reinforcement Learning. *IEEE Access, 10*, 35712-35724.

O'Neill, S., Bond, R. R., Grigorash, A., Ramsey, C., Armour, C., & Mulvenna, M. D. (2019). Data analytics of call log data to identify caller behaviour patterns from a mental health and well-being helpline. *Health Informatics Journal, 25*(4), 1722-1738.

Zhang, M., Sheng, Y., Tian, N., Liu, W., Wang, H., Zhu, L., & Xu, Q. (2021). Research and application of traffic forecasting in customer service center based on ARIMA model and LSTM neural network model. *Journal of Physics: Conference Series, 1881*(3), 032063

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.