

## การพยากรณ์ยอดขายของ Walmart โดยใช้ตัวแบบการเรียนรู้ของเครื่อง

### SALES FORECASTING OF WALMART USING MACHINE LEARNING MODEL

นลินทิพย์ ชื่นใจ<sup>1\*</sup>, นงนภัศ ดวงศรี<sup>1</sup>, กชกร สามารถ<sup>1</sup>, นริสา ทีเก่ง<sup>1</sup>, เดชาธร ปัทวิง<sup>1</sup>, เฉลิมรัช นนทะภา<sup>2</sup>

Nalinthip Chuenjai<sup>1\*</sup>, Nongnapat Duangsri<sup>1</sup>, Kodchakorn Samart<sup>1</sup>, Narisa Teekeng<sup>1</sup>, Dachatorn Pattawing<sup>1</sup>, Chalermrat Nontapa<sup>2</sup>

<sup>1</sup>โครงการจัดตั้งศูนย์วิจัยวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

<sup>2</sup>ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

\* ผู้บริหารประสานงาน: นลินทิพย์ ชื่นใจ

อีเมล: nalinthip\_c@cmu.ac.th

#### บทคัดย่อ:

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบการพยากรณ์ยอดขายของ Walmart และเปรียบเทียบตัวแบบการพยากรณ์ทั้ง 4 ตัวแบบ คือ Linear Regression, Random Forest, LightGBM และ XGBoost โดยใช้ข้อมูลรายสัปดาห์ของรายได้ของ Walmart ตั้งแต่เดือนกุมภาพันธ์ พ.ศ. 2553 ถึงเดือนตุลาคม พ.ศ. 2555 รวมทั้งสิ้น 32 เดือน หรือ 142 สัปดาห์ โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ข้อมูลชุดแรกเป็นข้อมูลฝึกสอนตั้งแต่เดือนกุมภาพันธ์ พ.ศ. 2553 ถึงเดือนมีนาคม พ.ศ. 2555 จำนวน 25 เดือน เพื่อใช้ในการสร้างตัวแบบในการพยากรณ์ และข้อมูลชุดที่สองเป็นข้อมูลชุดทดสอบตั้งแต่เดือนเมษายน พ.ศ. 2555 ถึงเดือนตุลาคม พ.ศ. 2555 จำนวน 7 เดือน เพื่อเปรียบเทียบค่าความถูกต้องโดยเกณฑ์ที่ใช้เปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ คือ ค่าเฉลี่ยสัมบูรณ์ของเปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) ค่าสัมประสิทธิ์การตัดสินใจ (R-Squared) ค่าเฉลี่ยของความคลาดเคลื่อนแบบค่าสัมบูรณ์ (MAE) รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) และค่าที่ใช้วัดความคลาดเคลื่อนของโมเดลแบบ log scale (RMSLE) ผลการวิจัยพบว่า ตัวแบบที่เหมาะสมที่สุด คือ ตัวแบบ Linear Regression โดยมีค่า MAPE เท่ากับ 4.92% มีค่า R-Squared เท่ากับ 98.56% ค่า MAE เท่ากับ 45,896.85 มีค่า RMSE เท่ากับ 64,116.15 และมีค่า RMSLE เท่ากับ 0.063

คำสำคัญ : การพยากรณ์ยอดขาย, การเรียนรู้ของเครื่อง, Walmart, Linear Regression, XGBoost

#### Abstract:

This research aimed to develop a sales forecasting model for Walmart and to compare the performance of four forecasting models: Linear Regression, Random Forest, LightGBM, and XGBoost. The study utilized Walmart's weekly revenue data from February 2010 to October 2012, covering a total period of 32 months or 142 weeks. The dataset was divided into two subsets. The first subset comprised training data from February 2010 to March 2012 or 25 months, which was used to develop the forecasting models. The second subset consisted of testing data from April 2012 to October 2012 (7 months), which was used to evaluate and compare the predictive accuracy of the models. The evaluation metrics included Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R-Squared), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Root Mean Squared Logarithmic Error (RMSLE). The results indicated that the Linear Regression model was the most suitable model, achieving a MAPE of 4.92%, an R-Squared value of 98.56%, an MAE of 45,896.85, an RMSE of 64,116.15, and an RMSLE of 0.063.

Keywords : Sales Forecasting, Machine Learning, Walmart, Linear Regression, XGBoost