SUPER AI ENGINEER

By CP Axtra AI Team

05.05.2025

Topic I
Sales Forecasting

Sales Forecasting at Store Item Level (in Baht)

To build a model that accurately forecasts the sales revenue (in Thai Baht) for each item in various retail stores.

The goal is to predict daily sales for the next month based on historical sales data, promotions, holidays, and other relevant factors. Encourage participants to add more data such as local event, rainfall, temperature, AQI, etc.
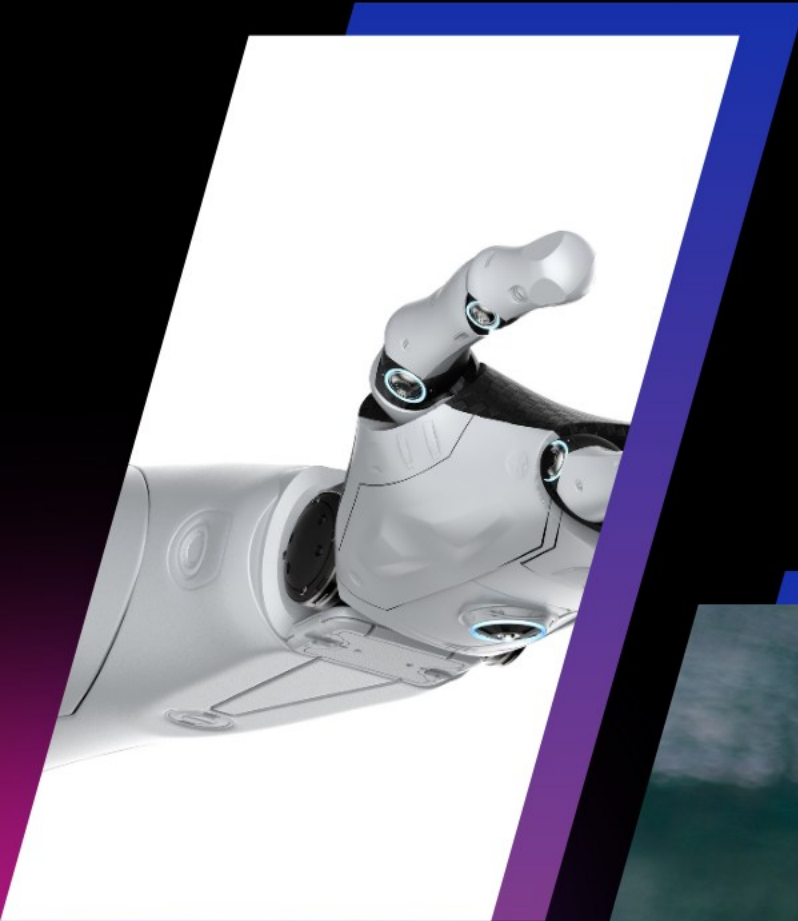
Evaluation Metrics

Weighted Absolute Percentage Error (WAPE) :

Measures the overall deviation of forecasted values from observed values

$$WAPE = \frac{\sum_{i,t} |y_{i,t} - \hat{y}_{i,t}|}{\sum_{i,t} |y_{i,t}|}$$

$y_{i,t}$ – the observed value at point (i,t)

$\hat{y}_{i,t}$ – the predicted value at point (i,t)

Data

Dataset consists of 2 years and 3 months recording

Historical Sales:

- ○ date: Date of the sales record
- ○ store_id: Unique identifier for the store
- ○ item_id: Unique identifier for the item
- ○ sales: Sales revenue in Baht

Calendar Data:

- ○ date: Date
- ○ holiday: Description of any special event

Submission File

- ○ date: Date of the sales record
- ○ store_id: Unique identifier for the store
- ○ item_id: Unique identifier for the item
- ○ sales: Sales revenue in Baht (Forecasted)

Store Data:

- ○ store_id: Unique identifier for the store
- ○ region: region of store e.g., Bkk, North
- ○ size: Size of the store e.g., m^2
- ○ format: type of store e.g., small, medium, large

Item Hierarchy Data:

- ○ item_id: Unique identifier for the item
- ○ category: Category of the item e.g. Fresh Food,
- ○ subcategory: Subcategory of the item e.g. Pork, Egg

CP AXTRA

Topic II
Optimize Product Assortment

Optimize Product Assortment (add/delete item)

in Seasonal Periods for Each Store

The goal is to identify which items should be added or removed from the assortment to maximize sales and customer satisfaction during specific seasons.

Evaluation Metrics

- F1-score: evaluates the balance between precision and recall in the classification, based on the status decision of each item (3-classes: add, delete, or remain).

Baseline criterion (definition of each class):

For example: considering for W10-2025, referenced from W6-2025

- Add: W10-2025 has sales but W6-2025 doesn't have sales

- Delete: W10-2025 doesn't have sales but W6-2025 has sales

- Remain: W10-2025 has sales and W6-2025 has sales

These are rules-based, considering only from the sales recording without considering the sales_unit

Data

Dataset consists of 2 years and 3 months recording

Historical Sales Data:

- ○ date: Date of the sales record
- ○ store_id: Unique identifier for the store
- ○ item_id: Unique identifier for the item
- ○ sales: Sales revenue in Baht

Item Hierarchy Data:

- ○ item_id: Unique identifier for the item
- ○ category: Category of the item
- ○ subcategory: Subcategory of the item
- ○ brand: Brand of the item

Store Data:

- ○ store_id: Unique identifier for the store
- ○ region: region of store e.g., BKK, North
- ○ size: Size of the store e.g., m^2
- ○ format: type of store e.g., small, medium, large

Calendar Data:

- ○ date: Date
- ○ week: week of year e.g. 202515
- ○ holiday: Description of any special event

Submission File

- ○ YEAR_WEEK_NUMBER: Evaluation period (Weekly)
- ○ store_id: Unique identifier for the store
- ○ item_id: Unique identifier for the item
- ○ ITEM_STATUS: Forecast of item status

("REMAIN", "ADD", "DELETE")

CP AXTRA

Competition rules on Kaggle

- External dataset is allowed (Public dataset only)

- Encourage to use related features

(such as local event, temperature AQI)

- This dataset will be used only in this competition

- API does not allowed.

- Team name: House name, e.g. Machima etc.
(ตั้งชื่อทีม เป็นชื่อบ้านภาษาอังกฤษ เท่านั้น)

- Submission person: Only one represented member of each house team member
(ผู้ส่งผลในแต่ละการแข่งขัน เป็นตัวแทนบ้าน 1 คน เท่านั้น)

- หากพบการส่งผลจากแต่ละบ้านมากกว่า 2 คน ต่อโจทย์การแข่งขัน ขอสงวนสิทธิ์ในการไม่พิจารณาร่วมการแข่งขัน

<u>Submission on Kaggle</u>

- Start: May 5, 2025 at 12.00
- End: May 9, 2025 at 8.00
- Submission quota: 4 submission/day/competition (Reset 7.00 am)
- Scored private submissions: 2/competition

Let's go !

Competition 1: Sales Forecasting
https://www.kaggle.com/t/326491a037f94927b5861c92d51fc469

Competition 2: Optimize Product Assortment
https://www.kaggle.com/t/cd13006b5e084272b8f1686c265f0ec6

Q&A