

# Kaggle — M5 Forecasting

Kaggle website: <https://www.kaggle.com/c/m5-forecasting-accuracy>  
<https://www.kaggle.com/c/m5-forecasting-uncertainty>

MAFS5440, Fall 2024

# Background

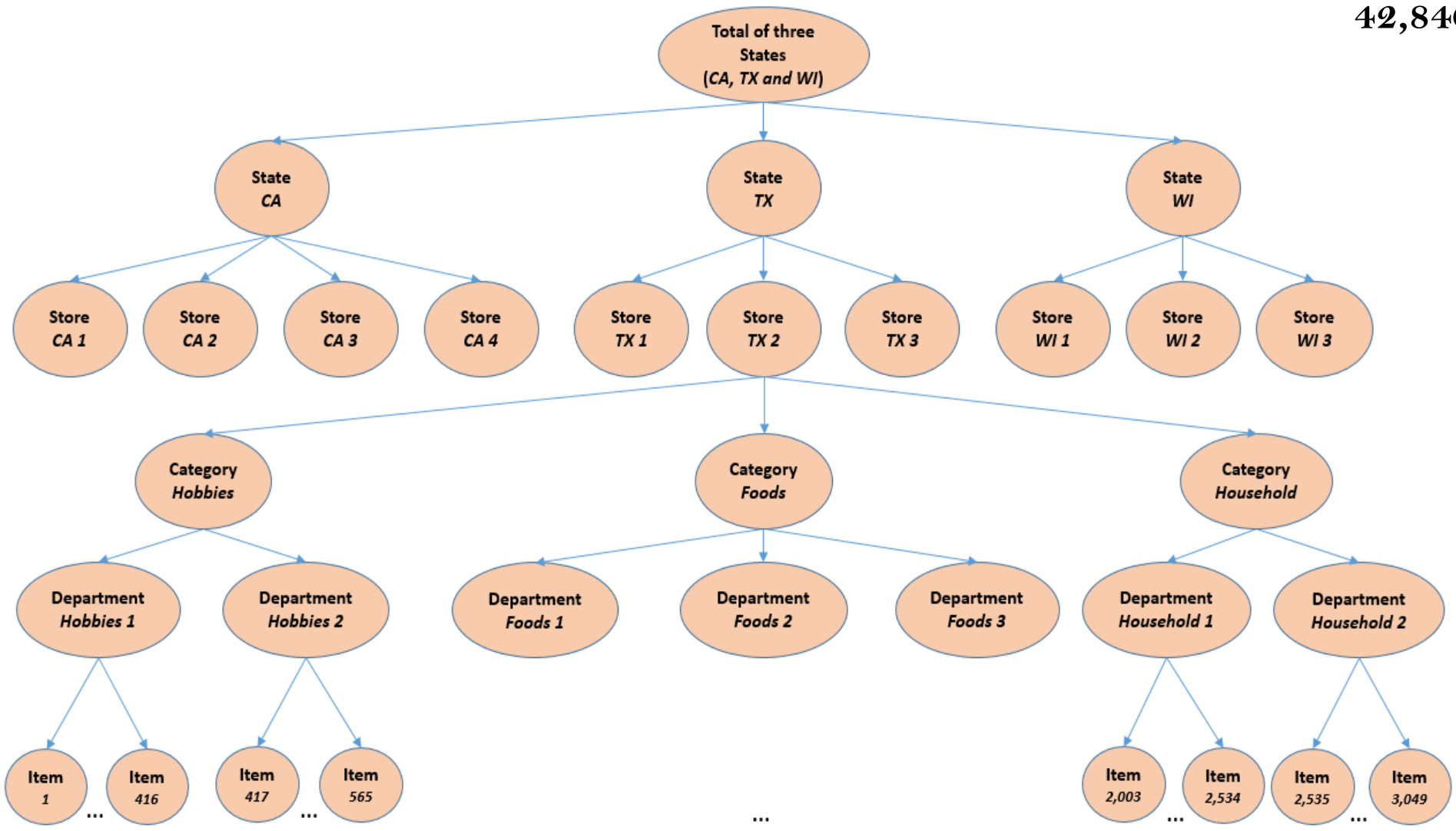
- M5 forecasting: the 5<sup>th</sup> Makridakis Competition.
- Task: Forecasting (**accuracy**) and estimating the **uncertainty** distribution of the realized values of the same series
  - **Accuracy task:** Can you estimate, as precisely as possible, the **point forecasts** of the unit sales of various products sold in the USA by Walmart?
  - **Uncertainty task:** Can you estimate, as precisely as possible, the **uncertainty distribution** of the unit sales of various products sold in the USA by Walmart?
- Aim:
  - Identifying the most appropriate method(s) for different types of situations requiring predictions and making uncertainty estimates
  - Comparing the accuracy/uncertainty of ML and DL methods versus those of standard statistical ones

# Data

- 42,840 time series data from Walmart (sales data from 2011-01-29 to 2016-06-19).
- **hierarchical sales data:** starting at the **item level** and aggregating to that of **departments, product categories** and **stores** in three geographical areas of the US: California, Texas, and Wisconsin.
- **explanatory variables** are also included; such as price, promotions, day of the week, and special events (e.g. Super Bowl, Valentine's Day, and Orthodox Easter) that affect sales which are used to improve forecasting accuracy.
- The majority of the more than 42,840 time series display **intermittency** (sporadic sales including zeros).

# Data Organization Overview

In total:  
42,840 time series



# Your Job

- **Accuracy task:** forecasting daily sales of each products for the next 28 days.  
([m5-forecasting-accuracy](#))
  - **Uncertainty task:** 28 days ahead probabilistic forecasts for the **median** and four **prediction intervals (PIs)** (50%, 67%, 95%, and 99%).  
([m5-forecasting-uncertainty](#))
- The two task using the same dataset.

# Evaluation metrics

- Accuracy task: **Weighted Root Mean Squared Scaled Error (RMSSE)**

$$RMSSE = \sqrt{\frac{1}{h} \frac{\sum_{t=n+1}^{n+h} (Y_t - \hat{Y}_t)^2}{\frac{1}{n-1} \sum_{t=2}^n (Y_t - Y_{t-1})^2}}$$

where  $Y_t$  is the actual future value of the examined time series at point  $t$ ,  $\hat{Y}_t$  the generated forecast,  $n$  the length of the training sample, and  $h$  the forecasting horizon.

$$WRMSSE = \sum_{i=1}^{42,840} w_i * RMSSE$$

where  $w_i$  is the weight of the  $i_{th}$  series of the competition. A lower WRMSSE score is better.

# Evaluation metrics

- **Uncertainty task: Weighted Scaled Pinball Loss (WSPL)**

$$\mathbf{SPL}(\mathbf{u}) = \frac{\frac{1}{h} \sum_{t=n+1}^{n+h} (Y_t - Q_t(u)) u \mathbf{1}\{Q_t(u) \leq Y_t\} + (Q_t(u) - Y_t)(1-u) \mathbf{1}\{Q_t(u) > Y_t\}}{\frac{1}{n-1} \sum_{t=2}^n |Y_t - Y_{t-1}|}$$

where  $Y_t$  is the actual future value of the examined time series at point  $t$ ,  $Q_t(u)$  the generated forecast for quantile  $u$ ,  $n$  the length of the training sample,  $h$  the forecasting horizon, and  $\mathbf{1}$  the indicator function.

- Given that forecasters will be asked to provide the **median**, and the 50%, 67%, 95%, and 99% **PIs**,  $u$  is set to  $u_1=0.005, u_2=0.025, u_3=0.165, u_4=0.25, u_5=0.5, u_6=0.75, u_7=0.835, u_8=0.975, \text{ and } u_9=0.995$ .

$$\mathbf{WSPL} = \sum_{i=1}^{42,840} w_i * \frac{1}{9} \sum_{j=1}^9 \mathbf{SPL}(u_j)$$

where  $w_i$  is the weight of the  $i_{th}$  series of the competition and  $u_j$  the  $j_{th}$  out of the examined quantiles. A lower WSPL score is better.

# Weighting

M5 involves the unit sales of various products of **different selling volumes and prices** that are organized in a **hierarchical** fashion. Therefore, you must provide accurate forecasts across all hierarchical levels, **especially for series of high importance**, i.e. for series that represent significant sales, measured in US dollars.

To that end, the forecasting errors computed for each participating method (both RMSSE and SPL) will be **weighted** across the M5 series based on their **cumulative actual dollar sales**, which is a good and objective proxy of their actual value for the company in monetary terms.

Refer: <https://github.com/Mcompetitions>