IMPROVING DEMAND FORECASTING

Lily Kou, Megan Moffatt, Shreya, Pankaj Nandal, Yoon Nam

05/04/25

Walmarts Demand Challenge

Demand at the item-store level is volatile and hard to predict — especially around promotions, holidays, and regional events.

Forecasting error directly

Impacts both cost and customer experience:

- Overstock → inventory waste
- Under-stock → missed sales and customer frustration

Previously,
Walmart

Previously, Walmart relied on **trend lines** or **store manager** input to guide inventory planning.

Previously,
Walmart

Previously, Walmart relied on **trend lines** or **store manager** input to guide inventory planning.

That approach **led to missed opportunities**, especially during **sales events** or **seasonal changes**.

Now,
Walmart

Now, Walmart can leverage our Machine Learning solution, predicting **sales across items, stores, departments, and states**.

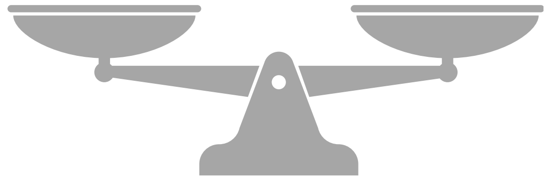
Now,
Walmart

Now, Walmart can leverage our Machine Learning solution, predicting **sales across items, stores, departments, and states**.

Our model achieves a **WRMSSE score of 0.61**, which signals strong performance.

Why Use WRMSSE?

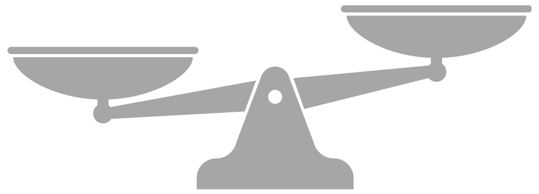
Not all forecasting errors cost the same.



Accuracy

Measures treat all errors the same

Predicting too high on lightbulbs counts just as much as predicting too low on TVs.



WRMSSE

Weighs errors differently

Mistakes on high-impact products like TVs or holiday staples are penalized more — because they cost Walmart more.

39% Reduction** in Forecast Error* | *WRMSSE Score = **0.61

Why 39% Matters

*At Walmart's scale, a **39% drop in error isn't small** —*
*it means **millions saved** in inventory decisions and **happier customers**.*

Naive Forecast

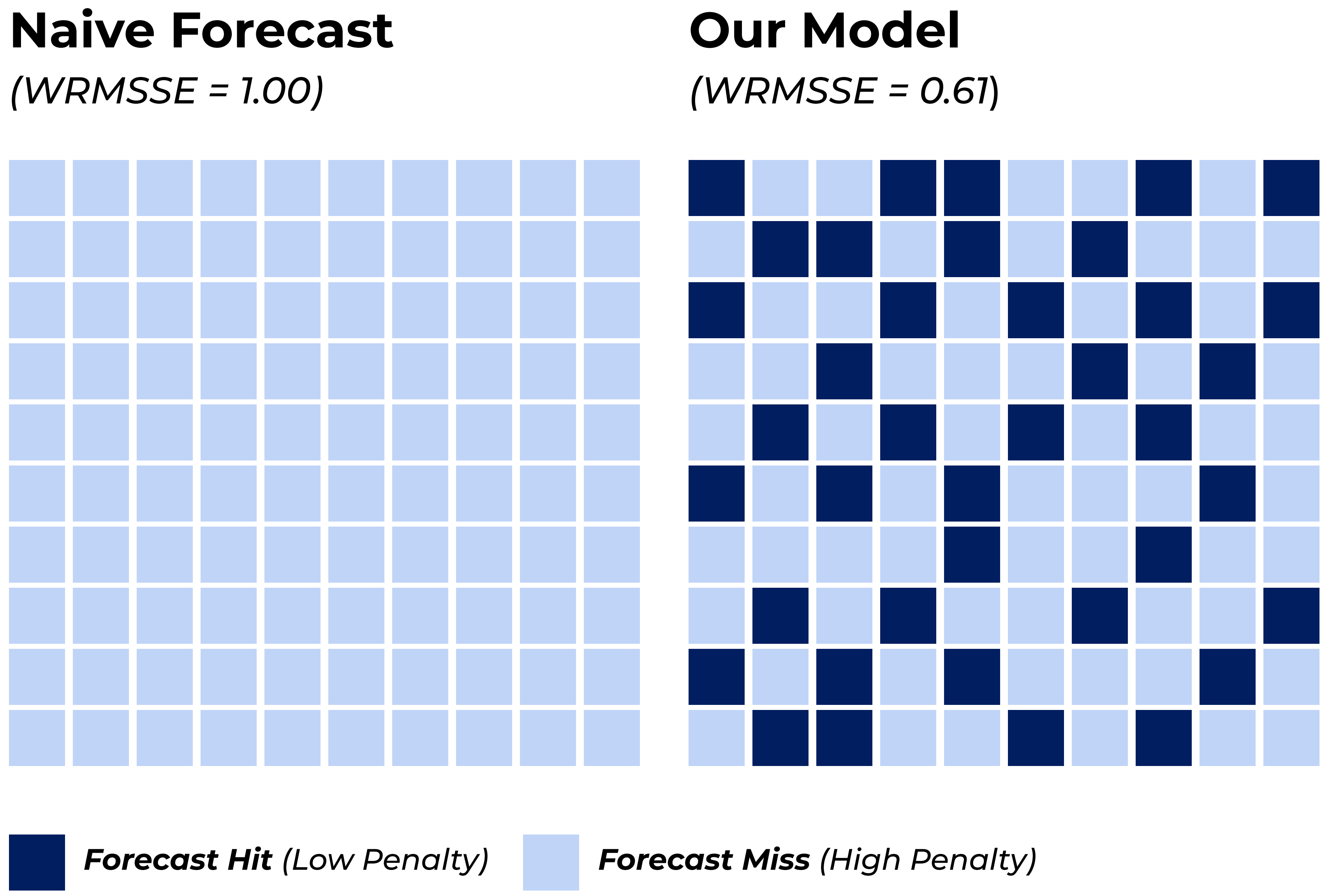
(WRMSSE = 1.00)

***Forecast Hit** (Low Penalty)* ***Forecast Miss** (High Penalty)*

39% Reduction** in Forecast Error* | *WRMSSE Score = **0.61

Why 39% Matters

*At Walmart's scale, a **39% drop in error isn't small** —
it means **millions saved** in inventory decisions and **happier customers**.*



How our Model Understands Data

Static

item_id, store_id, category

Basic characteristics that identify each item and store

Calendar

weekday, holidays, seasonality

Helps the model recognize patterns across days, weekends, months, holidays, and seasons.

How our Model Understands Data

Static

item_id, store_id, category

Basic characteristics that identify each item and store

Calendar

weekday, holidays, seasonality

Helps the model recognize patterns across days, weekends, months, holidays, and seasons.

Promotion

SNAP flags

Indicates active government SNAP benefits, helping the model capture subsidy-driven demand changes.

Lag + Rolling Patterns

autocorrelation, seasonality patterns

Captures past sales trends, seasonality, and momentum — helping forecast future behavior.

Models We Tested

Best performer: ***LightGBM at WRMSSE = 0.61***

Models We Tested

Best performer: *LightGBM at WRMSSE = 0.61*

Model Type	Why Try It?	Key Insight
Linear	Fast & interpretable baseline to benchmark performance	Performed better than expected **(WRMSSE = 0.88)**, but missed nonlinear effects

Models We Tested

Best performer: ***LightGBM at WRMSSE = 0.61***

Model Type	Why Try It?	Key Insight
Linear	Fast & interpretable baseline to benchmark performance	Performed better than expected **(WRMSSE = 0.88)**, but missed nonlinear effects
Tree-Based	Handles lags, categoricals, and nonlinear relationships	**LightGBM** outperformed all others **(WRMSSE = 0.61)**; strong with calendar and pricing

Models We Tested

Best performer: ***LightGBM at WRMSSE = 0.61***

Model Type	Why Try It?	Key Insight
Linear	Fast & interpretable baseline to benchmark performance	Performed better than expected **(WRMSSE = 0.88)**, but missed nonlinear effects
Tree-Based	Handles lags, categoricals, and nonlinear relationships	**LightGBM** outperformed all others **(WRMSSE = 0.61)**; strong with calendar and pricing
Neural Net	Captures time-based patterns and complex interaction	**LSTM models** captured sequences well but were slower and less performant **(WRMSSE ≈ 1.0)**

Models We Tested

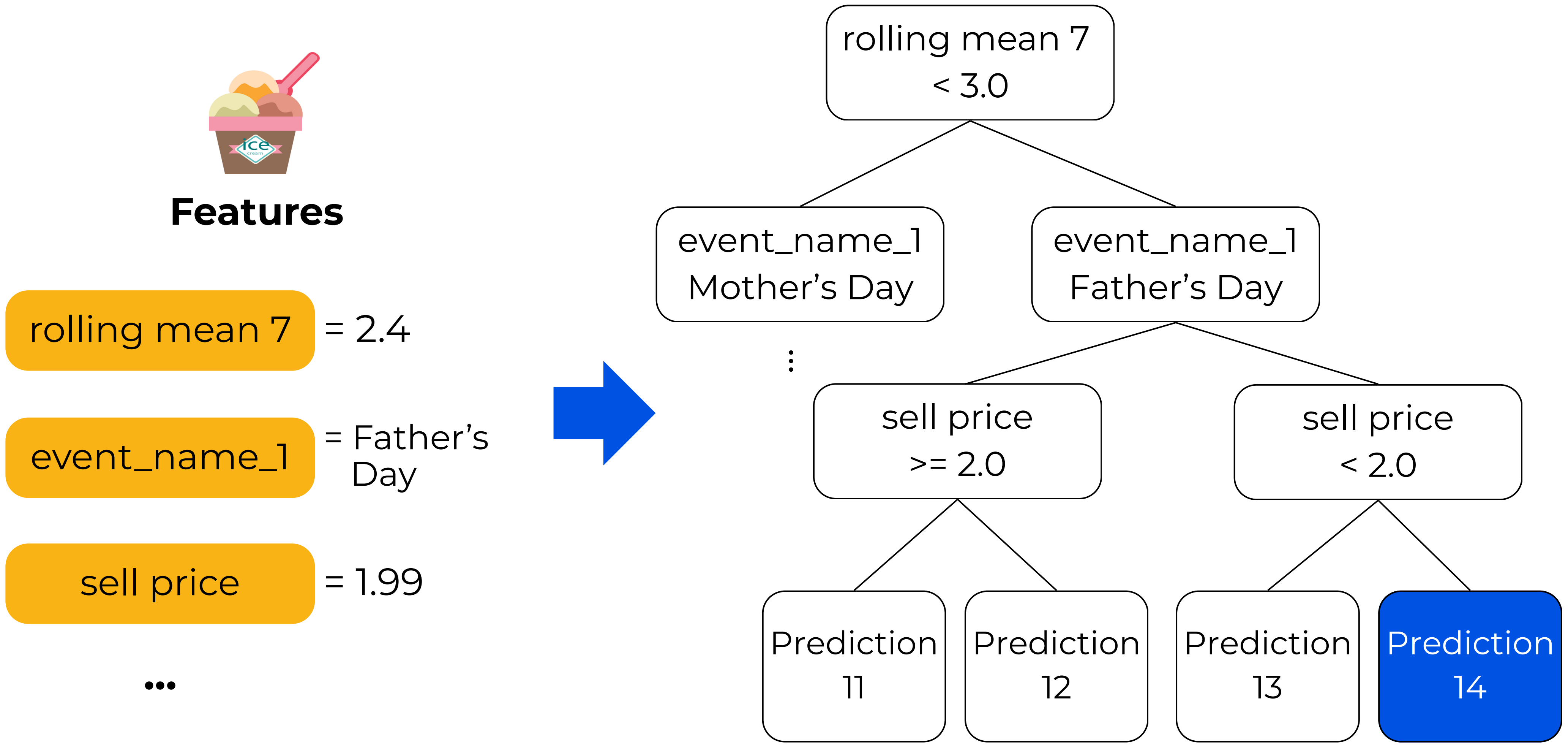
Best performer: ***LightGBM at WRMSSE = 0.61***

Model Type	Why Try It?	Key Insight
Linear	Fast & interpretable baseline to benchmark performance	Performed better than expected **(WRMSSE = 0.88)**, but missed nonlinear effects
Tree-Based	Handles lags, categoricals, and nonlinear relationships	**LightGBM** outperformed all others **(WRMSSE = 0.61)**; strong with calendar and pricing
Neural Net	Captures time-based patterns and complex interaction	**LSTM models** captured sequences well but were slower and less performant **(WRMSSE ≈ 1.0)**
Ensemble	Combines strengths across models for robustness	**Matched LightGBM**'s score but **added complexity** without significant lift

LightGBM

How does our model predict future sales?

Fast, accurate sales forecasts — by splitting where it counts.



What Does 39% Mean for Walmart's Bottom Line?

Better forecasts = better shelves, fewer markdowns, and millions in avoided costs.

- **Stockouts**

  Fewer stockouts on high-demand items during peak events

- **Inventory**

  Less excess inventory sitting idle

- **Ordering**

  Smarter ordering decisions at the store level

Limitations

Working with Real-World Data

30,000+

Unique item-store combinations

1,900+

Days of daily sales data

3,900+

Columns across 5 joined datasets

Recommendations

Next Steps

Deploy Model

Deploy the model through Walmart's cloud environment

Integration

Integrate it with store-level systems for automated replenishment

Expansion

Expand features with other impactful data.

Walmart

THANK YOU
QUESTIONS?

Lily Kou, Megan Moffatt, Shreya, Pankaj Nandal, Yoon Nam

05/04/25

Appendix