Purdue University
Daniels School of Business

Final Project Presentation

Machine Learning

Presented By
Aishwarya Ajaykumar
Devarshi Sharma

Agenda

- Problem Definition
- The M5 Walmart Data
- Sales Forecasting
- Uncertainty determination
- Promotion impact Analysis on Price
- Image Recognition on Groceries

Problem Definition

Solution # 1

Sales Forecasting: Enhance inventory management and optimize supply chain decisions by predicting weekly sales of camping gear for the next 28 days, allowing for effective stock allocation and reduced holding costs. The company's first proposed solution.

Solution # 2

Uncertainty Estimation: Improve demand planning accuracy by forecasting daily sales while quantifying prediction uncertainty, enabling businesses to strategize buffer stock levels and manage risk more effectively.

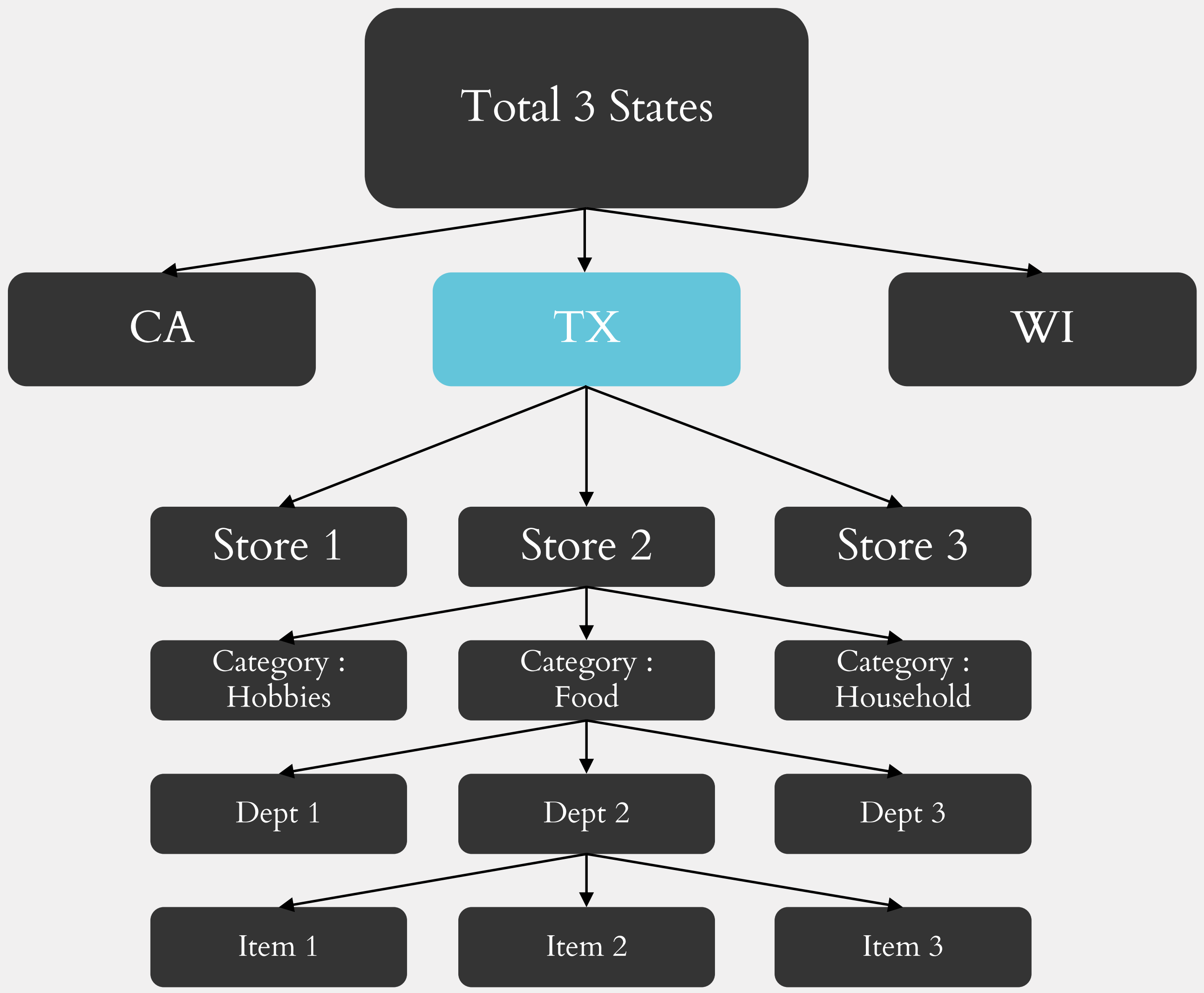
Solution # 3

Promotion and Price Impact Analysis: Drive revenue growth and optimize marketing spend by analysing the influence of promotional activities on price and sales, ensuring a high ROI on marketing campaigns.

Solution # 4

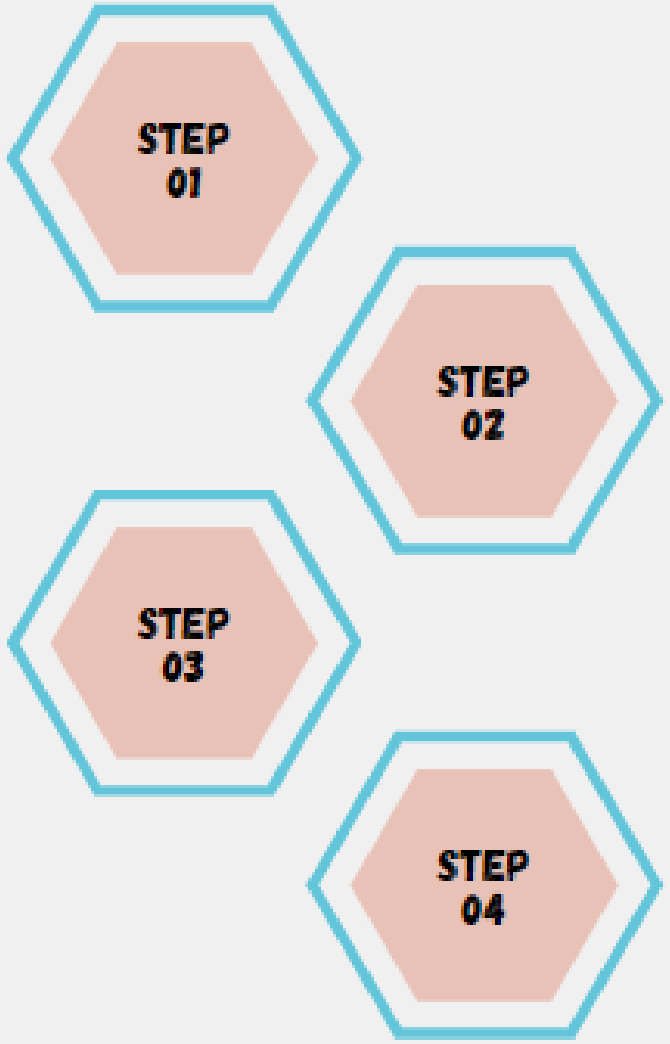
Walmart's use of image recognition technology can streamline the shopping experience by enhancing inventory management, expediting checkout processes, and improving customer service efficiency.

The Data



KEY OBSERVATIONS

- ❖ We have 6 years of data from 2011-2016(Only 4 months for 2016)
- ❖ Texas data has 3 stores, with 3 categories each, 7 departments each and a total of ~7600 items
- ❖ We are using Day 1 to 1914 as Train, Day 1914 to 1942 as Test



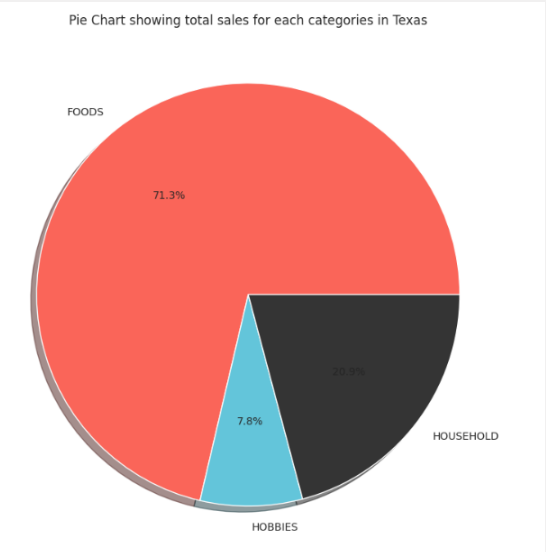
- ❖ All **Null values** are imputed so that model can algorithm to function correctly
- ❖ Filtering applied for **Texas data**
- ❖ Data **aggregated** to month level in some cases

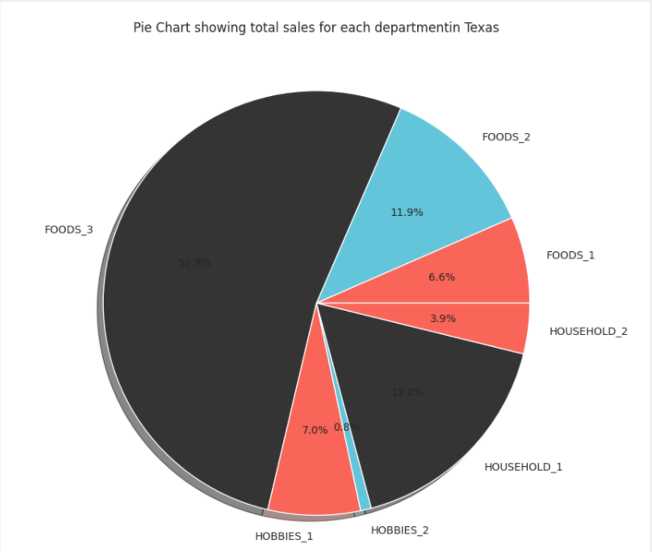
- ❖ Making all the **42840** combinations of the timeseries

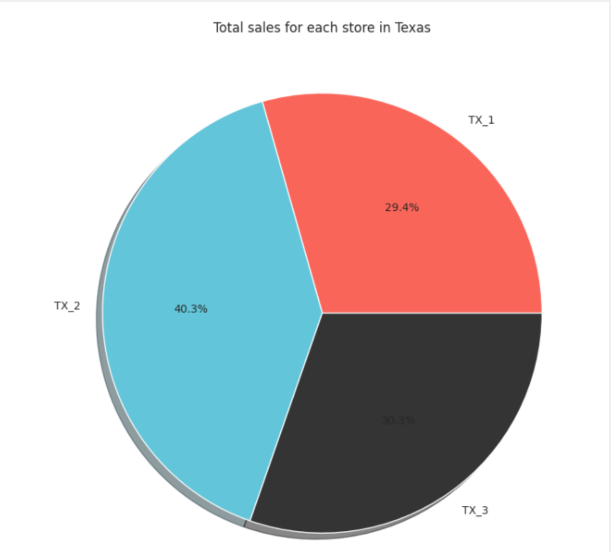
- ❖ **Season** information
- ❖ **Quarter, Month, Year** Start and end flags
- ❖ **Inflation** information
- ❖ General **holiday** information in the US added as supplementary information for prediction
- ❖ Data melted to make it in a long format in place of wide as depicted below

- ❖ **Lag features** to capture complex temporal patterns and improve forecasting performance.
- ❖ 4, 8, 16, 20, 24 weeks

	id	item_id	dept_id	cat_id	store_id	state_id	d	sales	date	wm_yr_wk	weekday	wday	month	year	event_name_1	event_type_1	event_name_2	event_type_2	snap_CA	snap_TX	snap_WI	sell_price
0	HOBBIES_1_008_CA_1_validation	HOBBIES_1_008	HOBBIES_1	HOBBIES	CA_1	CA	d_1	12	2011-01-29	11101	Saturday	1	1	2011	no_event	no_event	no_event	no_event	0	0	0	0.46
1	HOBBIES_1_008_CA_1_validation	HOBBIES_1_008	HOBBIES_1	HOBBIES	CA_1	CA	d_2	15	2011-01-30	11101	Sunday	2	1	2011	no_event	no_event	no_event	no_event	0	0	0	0.46
2	HOBBIES_1_008_CA_1_validation	HOBBIES_1_008	HOBBIES_1	HOBBIES	CA_1	CA	d_3	0	2011-01-31	11101	Monday	3	1	2011	no_event	no_event	no_event	no_event	0	0	0	0.46
3	HOBBIES_1_008_CA_1_validation	HOBBIES_1_008	HOBBIES_1	HOBBIES	CA_1	CA	d_4	0	2011-02-01	11101	Tuesday	4	2	2011	no_event	no_event	no_event	no_event	1	1	0	0.46
4	HOBBIES_1_008_CA_1_validation	HOBBIES_1_008	HOBBIES_1	HOBBIES	CA_1	CA	d_5	0	2011-02-02	11101	Wednesday	5	2	2011	no_event	no_event	no_event	no_event	1	0	1	0.46

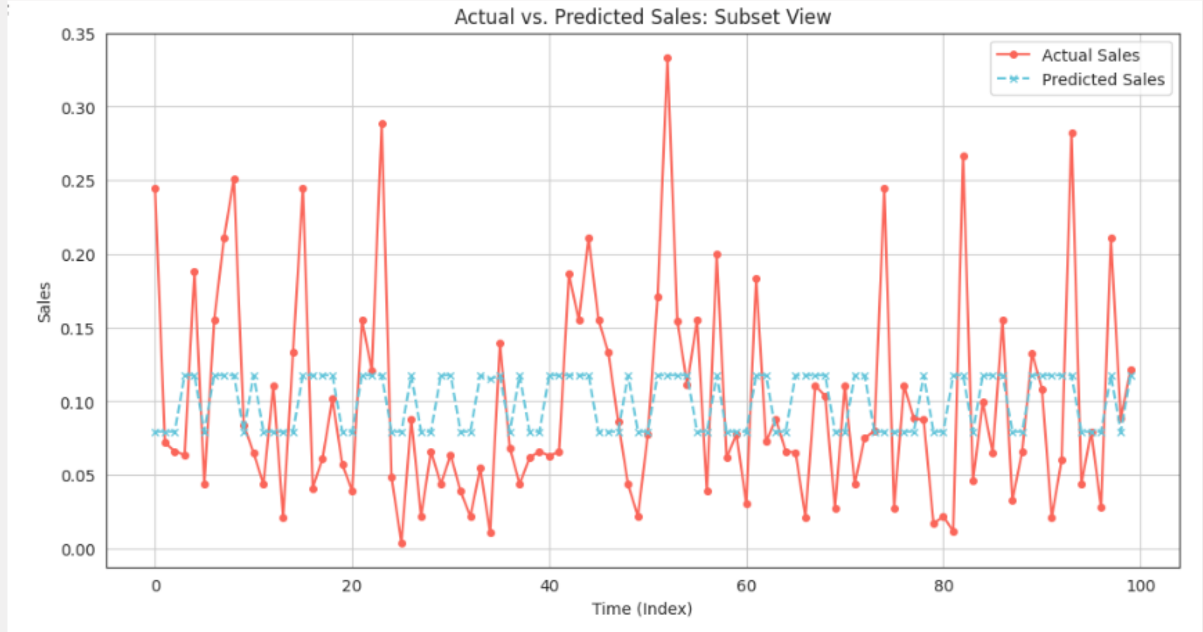


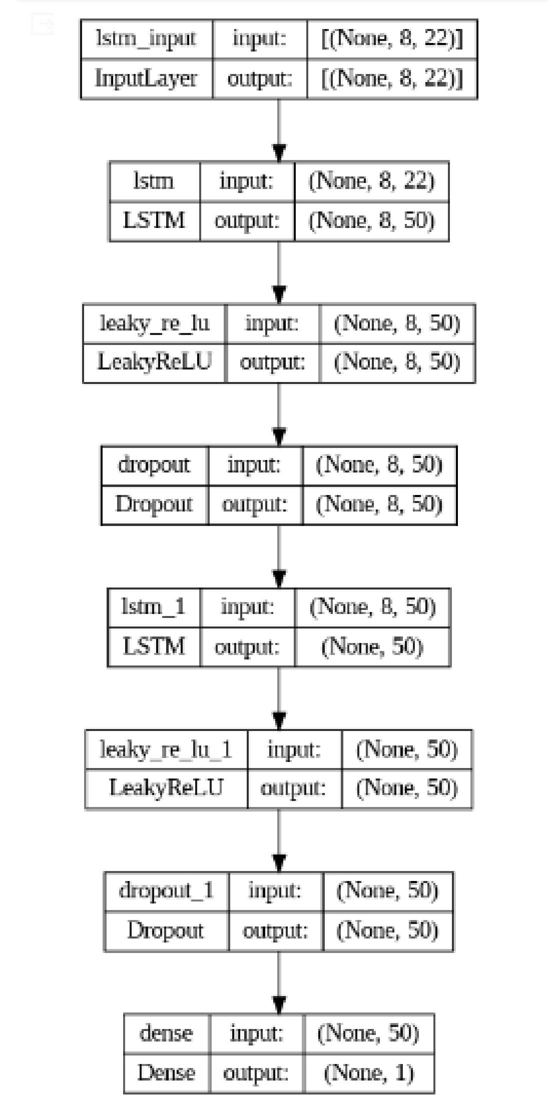




Sales Forecasting

```
Epoch 1/50
31113/31113 [==============================] - 166s 5ms/step - loss: 0.0052 - val_loss: 0.0051
Epoch 2/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0048 - val_loss: 0.0046
Epoch 3/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0048 - val_loss: 0.0047
Epoch 4/50
31113/31113 [==============================] - 158s 5ms/step - loss: 0.0049 - val_loss: 0.0047
Epoch 5/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0048 - val_loss: 0.0048
Epoch 6/50
31113/31113 [==============================] - 161s 5ms/step - loss: 0.0049 - val_loss: 0.0047
Epoch 7/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0048 - val_loss: 0.0048
Epoch 8/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0049 - val_loss: 0.0046
Epoch 9/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0048 - val_loss: 0.0046
Epoch 10/50
31113/31113 [==============================] - 158s 5ms/step - loss: 0.0048 - val_loss: 0.0049
Epoch 11/50
31113/31113 [==============================] - 160s 5ms/step - loss: 0.0048 - val_loss: 0.0050
Epoch 12/50
31113/31113 [==============================] - 161s 5ms/step - loss: 0.0048 - val_loss: 0.0051
8643/8643 [==============================] - 19s 2ms/step
RMSE: 0.06752067196897885
```





Data Preprocessing
- Normalization
- Sequence Length
- Feature Selection

Model Architecture
- Number of layers
- Number of Neurons
- Bidirectional LSTM

Regularization and Dropout
- Drop-out
- L2 Regularization

Training Strategies
- Batch Size and Epochs
- Learning Rate
- Early Stopping

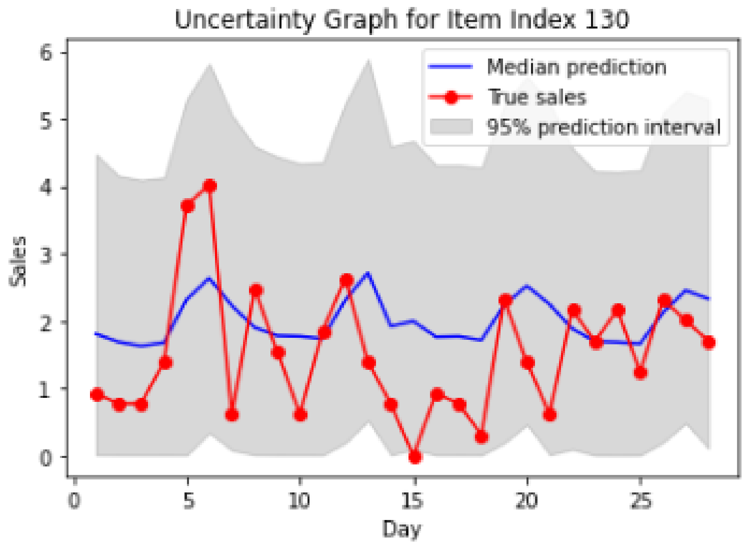
Inventory Optimization:
Sales forecasting in the supply chain enables businesses to anticipate demand fluctuations, aiding in the optimization of inventory levels to prevent overstocking or stockouts.

Improved Customer Service:
Anticipating sales trends allows businesses to better meet customer demand, ensuring products are available when needed, thereby enhancing customer satisfaction and loyalty.

Resource Allocation:
By accurately predicting sales, supply chain managers can efficiently allocate resources such as manpower, production capacity, and transportation, optimizing operations and reducing costs.

*Supply chains need at least a 4 week head-start to have adequate stock

Uncertainty Determination




```
Epoch 5/15
816/816 [==============================] - ETA: 0s - loss: 0.1978
Epoch 5: val_loss improved from 0.22296 to 0.22195, saving model to w.h5
816/816 [==============================] - 84s 103ms/step - loss: 0.1978 - val_loss: 0.2219 - lr: 0.0010
Epoch 6/15
816/816 [==============================] - ETA: 0s - loss: 0.1967
Epoch 6: val_loss improved from 0.22195 to 0.22118, saving model to w.h5
816/816 [==============================] - 89s 109ms/step - loss: 0.1967 - val_loss: 0.2212 - lr: 0.0010
Epoch 7/15
816/816 [==============================] - ETA: 0s - loss: 0.1958
Epoch 7: val_loss improved from 0.22118 to 0.22050, saving model to w.h5
816/816 [==============================] - 88s 108ms/step - loss: 0.1958 - val_loss: 0.2205 - lr: 0.0010
Epoch 8/15
816/816 [==============================] - ETA: 0s - loss: 0.1951
Epoch 8: val_loss improved from 0.22050 to 0.22023, saving model to w.h5
816/816 [==============================] - 86s 105ms/step - loss: 0.1951 - val_loss: 0.2202 - lr: 0.0010
Epoch 9/15
816/816 [==============================] - ETA: 0s - loss: 0.1943
Epoch 9: val_loss improved from 0.22023 to 0.21952, saving model to w.h5
816/816 [==============================] - 85s 104ms/step - loss: 0.1943 - val_loss: 0.2195 - lr: 0.0010
Epoch 10/15
816/816 [==============================] - ETA: 0s - loss: 0.1937
Epoch 10: val_loss improved from 0.21952 to 0.21914, saving model to w.h5
816/816 [==============================] - 89s 109ms/step - loss: 0.1937 - val_loss: 0.2191 - lr: 0.0010
Epoch 11/15
816/816 [==============================] - ETA: 0s - loss: 0.1932
Epoch 11: val_loss improved from 0.21914 to 0.21883, saving model to w.h5
816/816 [==============================] - 88s 108ms/step - loss: 0.1932 - val_loss: 0.2188 - lr: 0.0010
Epoch 12/15
816/816 [==============================] - ETA: 0s - loss: 0.1927
Epoch 12: val_loss improved from 0.21883 to 0.21853, saving model to w.h5
816/816 [==============================] - 87s 107ms/step - loss: 0.1927 - val_loss: 0.2185 - lr: 0.0010
Epoch 13/15
816/816 [==============================] - ETA: 0s - loss: 0.1923
Epoch 13: val_loss improved from 0.21853 to 0.21828, saving model to w.h5
816/816 [==============================] - 87s 106ms/step - loss: 0.1923 - val_loss: 0.2183 - lr: 0.0010
Epoch 14/15
816/816 [==============================] - ETA: 0s - loss: 0.1920
Epoch 14: val_loss improved from 0.21828 to 0.21822, saving model to w.h5
816/816 [==============================] - 86s 105ms/step - loss: 0.1920 - val_loss: 0.2182 - lr: 0.0010
```

Aggregation and Embedding of Features

Feature	Description
Weekday	Embedding for the weekday (`wday`) input.
Month	Embedding for the month input.
Year	Embedding for the year input.
Event	Embedding for the event name (`event`) input.
Day of Month	Embedding for the day of the month (`mday`) input.
Item ID	Embedding for the item ID (`item`) input.
Department ID	Embedding for the department ID (`dept`) input.
Category ID	Embedding for the category ID (`cat`) input.
Store ID	Embedding for the store ID (`store`) input.
State ID	Embedding for the state ID (`state`) input.

Flatten and Concatenate

Relu → **2 Dense Layered NN (125 each)**

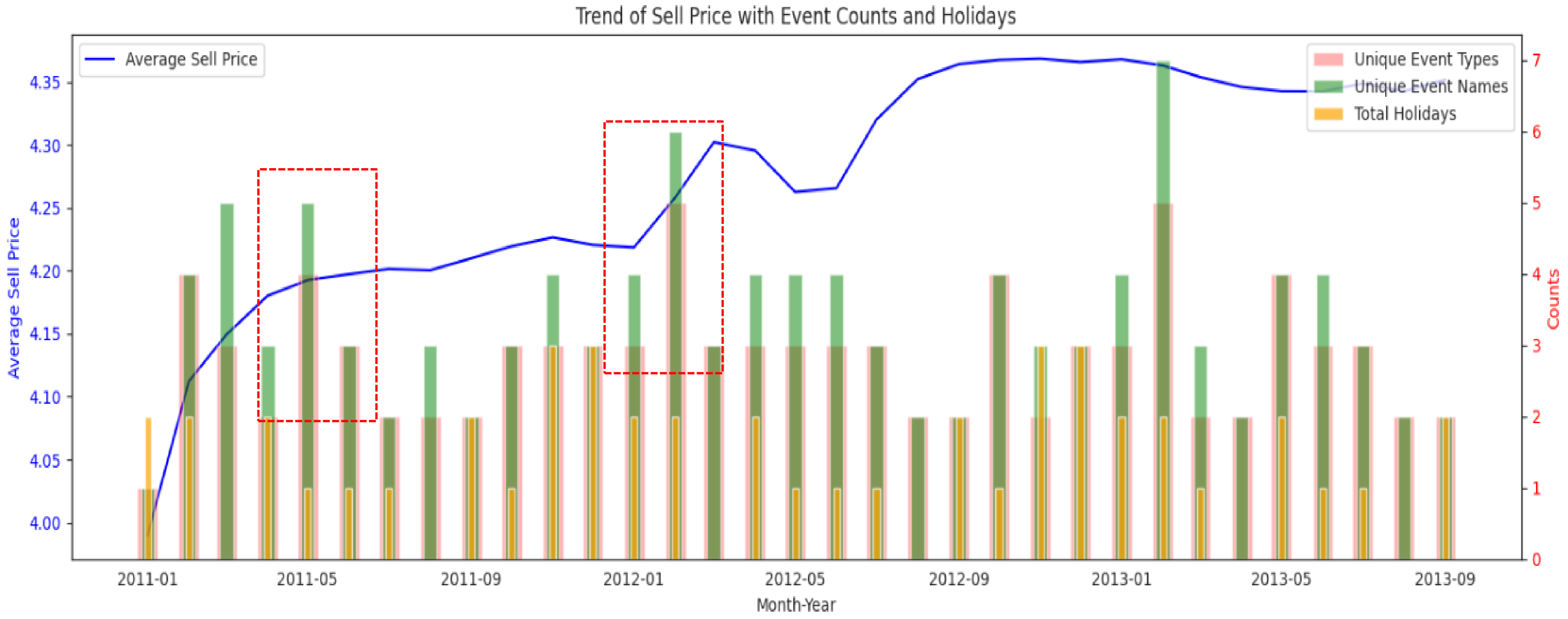
Linear → **Dense Layer (9 dimension)**

Risk Mitigation:
Uncertainty estimation in the supply chain allows businesses to identify and assess potential risks more accurately, enabling proactive risk management strategies to mitigate the impact of unforeseen events such as supply chain disruptions or demand fluctuations.

Optimized Inventory Management:
By quantifying prediction uncertainty, businesses can adjust buffer stock levels more precisely, ensuring adequate inventory to meet demand variability while minimizing excess inventory holding costs.

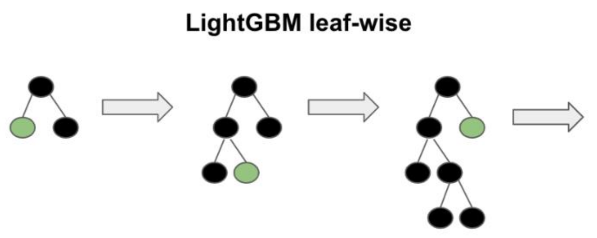
Enhanced Decision-Making:
Incorporating uncertainty estimation into demand planning processes provides decision-makers with more reliable forecasts and a clearer understanding of potential deviations, facilitating informed decision-making and improving overall supply chain resilience.

Promotion Impact Analysis



❖Correlation: There seems to be some correlation between events and sell prices, especially noticeable in periods where higher counts of events or holidays coincide with peaks in sell prices. This suggests that there might be a relationship between promotional activities, holidays, and pricing strategies.

❖Seasonality: The graph may also hint at seasonality in the data, where certain times of the year show higher prices or more events, possibly due to seasonal demand or seasonal promotional activities.




```
params = {
    'objective': 'regression',
    'metric': 'rmse',
    'num_leaves': 31,
    'learning_rate': 0.05,
    'verbose': -1  # This is to control the verbosity of the output
}
```

MAPE: 96.96178217224882%

	Feature	Importance	normalized_importance_percentage
0	event_name_1	7456.411274	45.974661
3	count_holidays	3903.221445	24.066441
2	snap_TX	2745.595858	16.928765
1	event_type_1	2113.295206	13.030133

Demand Forecasting and Inventory Management:
Promotion impact analysis facilitates **optimized inventory levels** by forecasting demand fluctuations during promotions and minimizing excess stock during non-promotional periods, enhancing overall inventory management efficiency.

Price Elasticity Estimation:
Analysing promotion effects aids in estimating price elasticity, guiding pricing strategies for maximizing revenue by understanding consumer responses to price changes.

Consumer Behaviour Insights:
Analysing the impact of promotions on pricing provides valuable insights into consumer behaviour, thereby enabling the **development of targeted marketing strategies** for improved sales and profitability.

*Assumption : Holiday events have a promotion associated with it

Image Recognition for Groceries

```
Epoch 16/50
16/16 [==============================] - 12s 804ms/step - loss: 0.1853 - accuracy: 0.9585 - val_loss: 0.1848 - val_accuracy: 0.9271
Epoch 17/50
16/16 [==============================] - 12s 752ms/step - loss: 0.1353 - accuracy: 0.9564 - val_loss: 0.1026 - val_accuracy: 0.9479
3/3 [==============================] - 1s 416ms/step - loss: 0.1026 - accuracy: 0.9479
Validation Loss: 0.10263439267873764
Validation Accuracy: 0.9479166865348816
```

```
[[ 3  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  4  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  6  0  0  0  0  0  0  0  0  1  0  0  0]
 [ 0  0  0  3  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  8  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  5  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  5  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  3  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  4  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  5  0  0  1  0  0]
 [ 0  1  0  0  0  0  0  0  0  0 20  0  0  0  1]
 [ 0  0  0  0  0  0  0  0  0  0  0 15  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  3  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0 25  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  6]]
```

```
4/4 [==============================] - 3s 357ms/step
```

	precision	recall	f1-score	support
Asparagus	1.00	1.00	1.00	3
Aubergine	0.80	1.00	0.89	4
Brown-Cap-Mushroom	0.86	0.86	0.86	7
Cabbage	1.00	1.00	1.00	3
Carrots	1.00	1.00	1.00	8
Cucumber	1.00	1.00	1.00	5
Garlic	1.00	1.00	1.00	5
Ginger	1.00	1.00	1.00	3
Leek	1.00	1.00	1.00	4
Onion	1.00	0.71	0.83	7
Pepper	1.00	0.91	0.95	22
Potato	0.94	1.00	0.97	15
Red-Beet	0.75	1.00	0.86	3
Tomato	1.00	1.00	1.00	25
Zucchini	0.86	1.00	0.92	6
accuracy			0.96	120
macro avg	0.95	0.97	0.95	120
weighted avg	0.96	0.96	0.96	120

Image Preparation

Model Architecture

Base Model Loading: Load MobileNetV2

Adding Layers: Append GlobalAveragePooling2D, Dense, Dropout, and a final Dense layer

Loss Function, Optimizer

Training Strategies

Class Weights

Callbacks: Use EarlyStopping and ModelCheckpoint

Enhanced Inventory Management: Image recognition technology can automate the tracking of stock levels, reducing manual inventory checks and improving accuracy. This leads to more efficient restocking processes and minimized out-of-stock scenarios, enhancing customer satisfaction.

Improved Checkout Efficiency: Utilizing image recognition at checkout can speed up the scanning process of groceries, reducing wait times for customers. This technology can identify products without traditional barcodes, streamlining the checkout process and improving overall customer experience.

Fraud Detection and Prevention: Image recognition can help in identifying discrepancies between the item scanned and the item billed, reducing instances of fraud. By ensuring that the item being scanned matches the product description and price, Walmart can enhance security and prevent losses due to pricing errors or fraudulent activities.

Thank You!