Amazon Forecast

Michael Lin
Sr. Solutions Architect
Amazon Web Services

The power of forecasting

IMPROVING BUSINESS OUTCOMES WITH MACHINE LEARNING

INVENTORY PLANNING

Improve demand planning at granular levels

Reduce waste, increase inventory turns, and improve in-stock availability

WORKFORCE PLANNING

More effectively staff to meet varying demand levels

Improve utilization, time to serve, and customer satisfaction

CAPACITY PLANNING

Make longer term decisions with more confidence

Improve capital utilization

FINANCIAL PLANNING

Plan for sales and top-line revenue

Effectively manage cash flows

The current landscape

Customers increasingly want more variety, they want it immediately, and they want it cheaply

Accurate and flexible forecasts are key; however, traditional methods cannot capture the increasing number of demand signals and complexity

Market leaders are investing in ML-driven forecasting to more effectively meet demand

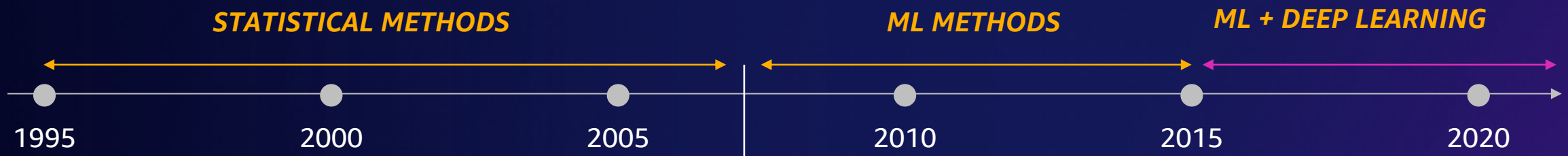
20% accuracy improvements

5% inventory reduction

3% increase in revenue

Harvard Business Review, McKinsey Global Institute
https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/most-of-ais-business-uses-will-be-in-two-areas

Amazon’s evolutionary forecasting journey



Best for few or independent data (e.g., aggregate demand)

- Aggregate demand
- Everyday household products
- Basic trends
- Repeated seasonal patterns

Best for complex and inter-related data (e.g., thousands of SKUs at hundreds of stores)

- Regional vs national demand
- Subtle seasonal patterns
- Relationships between products
- New products
- Slow moving products
- Changing prices or promotions
- External events

Leverage LCNC ML to optimize service levels

Amazon Forecast

Highly accurate Train and deploy ML models which can be up to 50% more accurate than traditional methods. No machine learning experience required

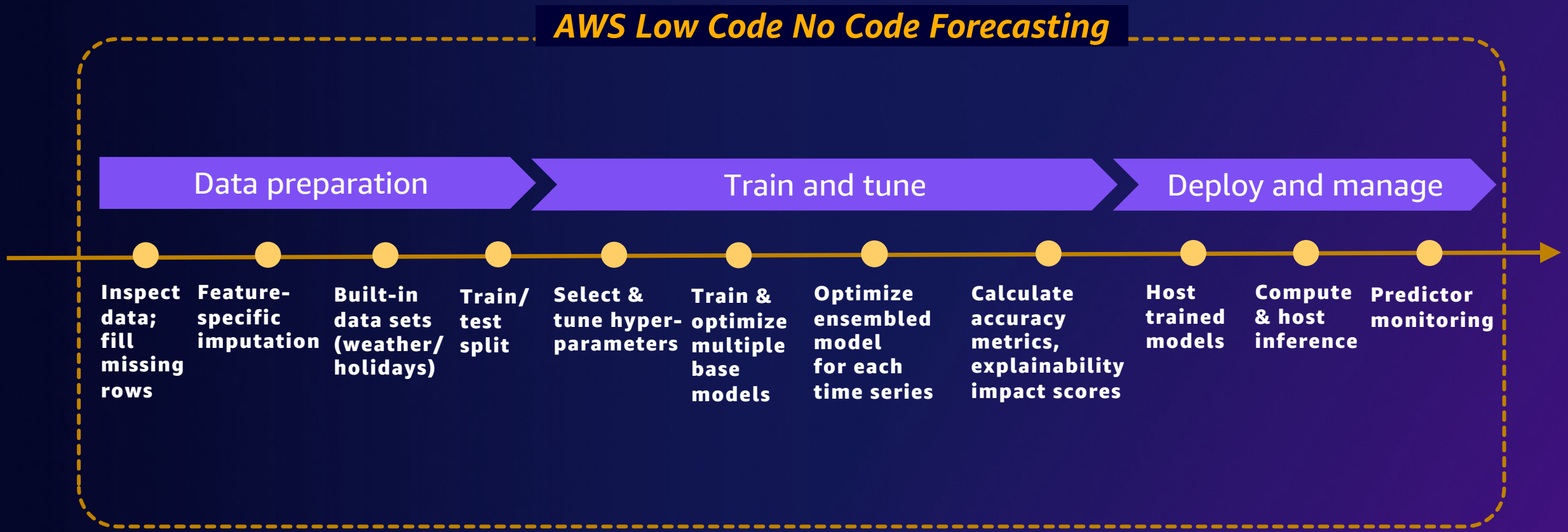
No code and no machine learning experience required. Leverage all the heavy lifting of building, training, and deploying custom models at scale

Easy to integrate into existing data lake, inventory, ordering, and supply chain systems

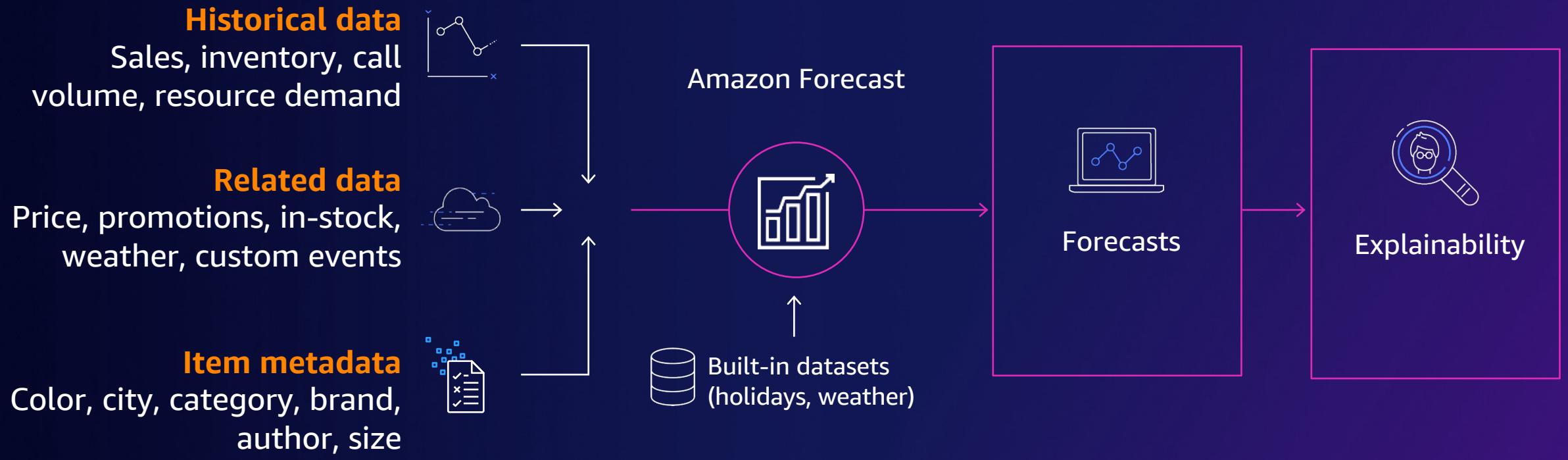
Quickly iterate, explore, prepare data, and onboard for ML-based forecasting. No need to bring entire data architecture onto AWS.

Fully managed ML forecasting

PREPARE, BUILD, TRAIN, TUNE, DEPLOY, AND MANAGE



Forecasting inputs and outputs



Data input examples

Target time series
(Historical demand)

Timestamp	*Item Id*	*Location*	*Target value*
Jan 1	1111	US_98121	50
Jan 1	2121	US_98121	89
Jan 2	1111	US_98003	35
Jan 2	3434	US_98003	73
Jan 3	1111	US_98003	45
Jan 3	5000	US_43505	13

Demand, sales, call volume, energy consumption, cloud storage consumption

Related time series

Timestamp	*Item Id*	*Location*	*Price*
Jan 1	1111	US_98121	$5
Jan 1	2121	US_98121	$20
Jan 2	1111	US_98003	$5
Jan 2	3434	US_98003	$13
Jan 3	1111	US_98003	$5
Jan 3	5000	US_43505	$4

Price, promotion, events, store hours, competitive average price

Item metadata

Item Id	*Color*	*Category*
1111	Yellow	Outdoor
2121	Red	Outdoor
3434	Green	Indoor
5000	Blue	Indoor

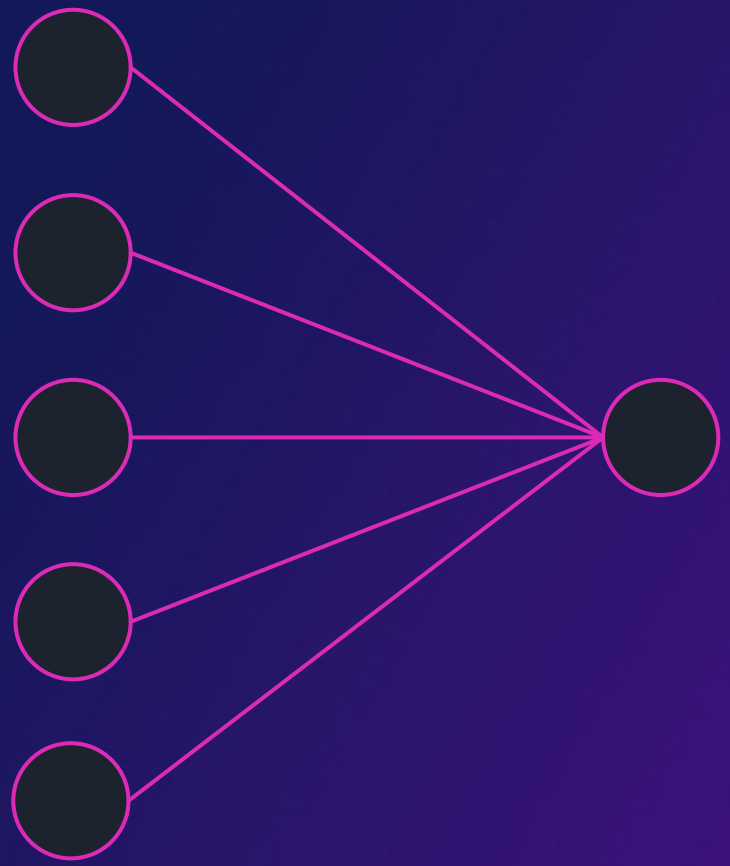
Category, department, color, texture

Tips:

1. Identify data sources and validate data is time series.
2. Gather large number of item IDs and historic data points (e.g. 3 years of daily data).
3. Clean and prepare data in consistent time stamp and frequency.

Training a model with AutoPredictor

- Up to 40% more accurate forecasts than before
- Chooses best models to ensemble at time-series level
- Incremental re-training: ~50% less time
- Model explanations
- Seamless upgrade from legacy predictors



Ensemble model algorithms to improve accuracy

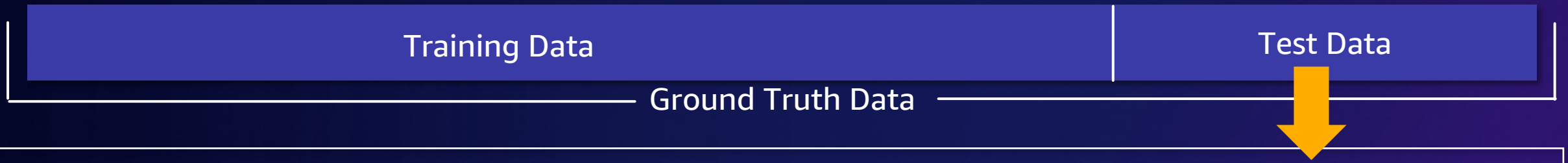
A GALLERY OF STATISTICAL AND MACHINE LEARNING FORECASTING ALGORITHMS

- Different products (SKUs / time-series) will exhibit different demand characteristics
- Different algorithms have their strengths
- Ensemble Modeling selects the best combination of algorithms for each product
- A more accurate forecast than any one single algorithm

Neural Networks		*Statistical Algorithms*			
CNN-QR	***DeepAR+***	***Prophet***	***NPTS***	***ARIMA***	***ETS***
Uses causal convolutional neural networks (CNNs).	Global neural model	Additive model with Gaussian likelihood	Non-parametric time series	Auto-regressive integrated moving average	Statistical algorithm that uses exponential smoothing
Works best with large datasets containing hundreds of time series; accepts related time series data without future values.	Uses related time series and attributes to train a model	Can find trend, seasonality, cyclical, and holiday effects	Performs well for intermittent spikes	Works well with a small number of time series; Classical approach to model autocorrelations	Works well with a small number of time series Finds trends, seasonality, and residual

Evaluating the quality of a predictor

TYPICAL ACCURACY METRICS FOR AN ML MODEL BASED ON ACTUAL GROUND TRUTH DATA

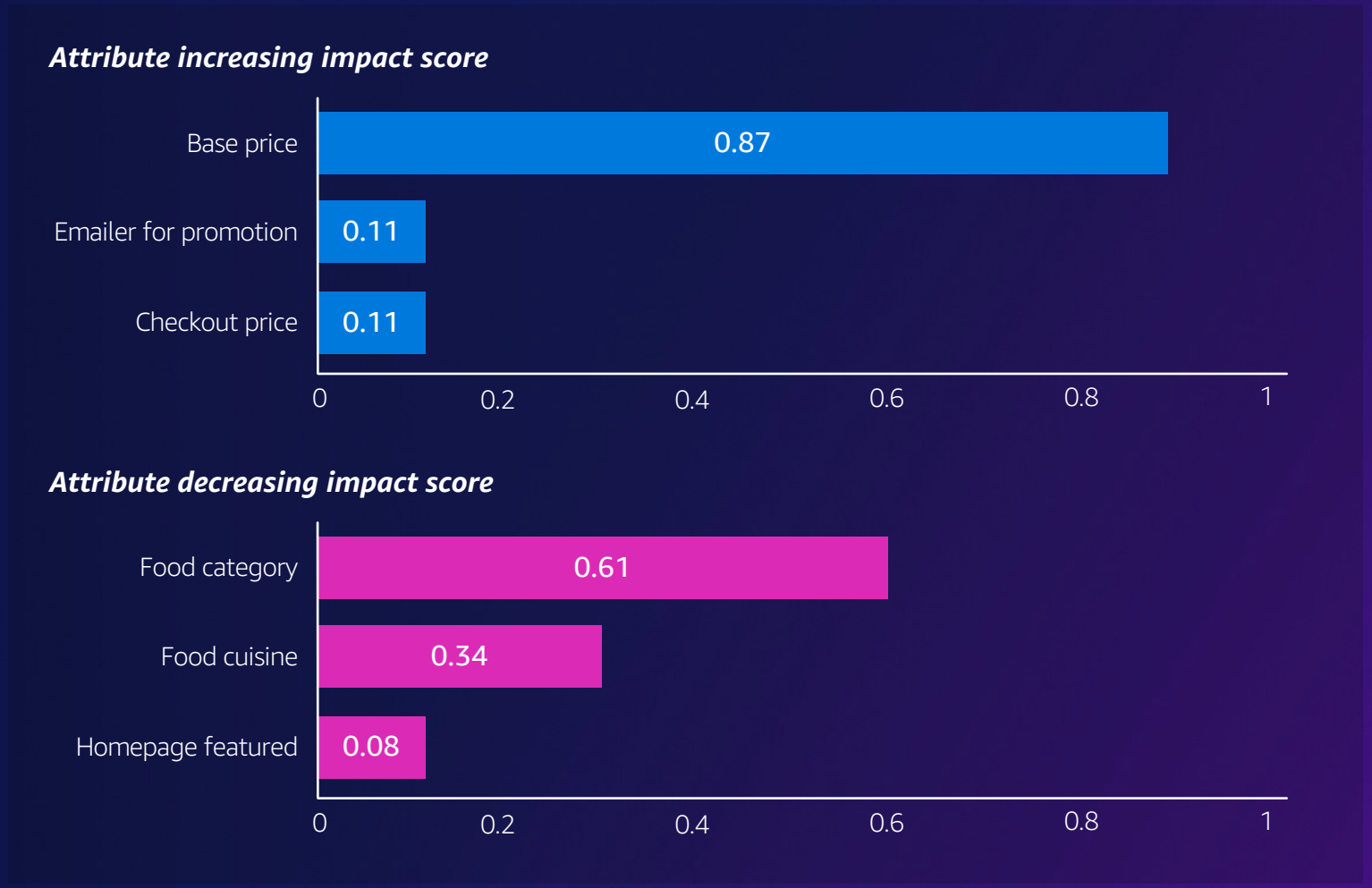


- Weighted Quantile Loss (wQL) - measures the accuracy of a model at a specified quantile. It is particularly useful when there are different costs for underpredicting and overpredicting.
- Weighted Absolute Percentage Error (WAPE) – measures the overall deviation of forecasted values from observed values.
- Root Mean Square Error (RMSE) – square root of the average of squared errors, and is therefore more sensitive to outliers than other accuracy metrics.
- Mean Absolute Percentage Error (MAPE) – absolute value of the percentage error between observed and predicted values for each unit of time, then averages those values.
- Mean Absolute Scaled Error (MASE) – divides the average error by a scaling factor. This scaling factor is dependent on the seasonality value, *m*, which is selected based on the forecast frequency
- Build your own, many variants of the above!

Understand your forecasts with explainability

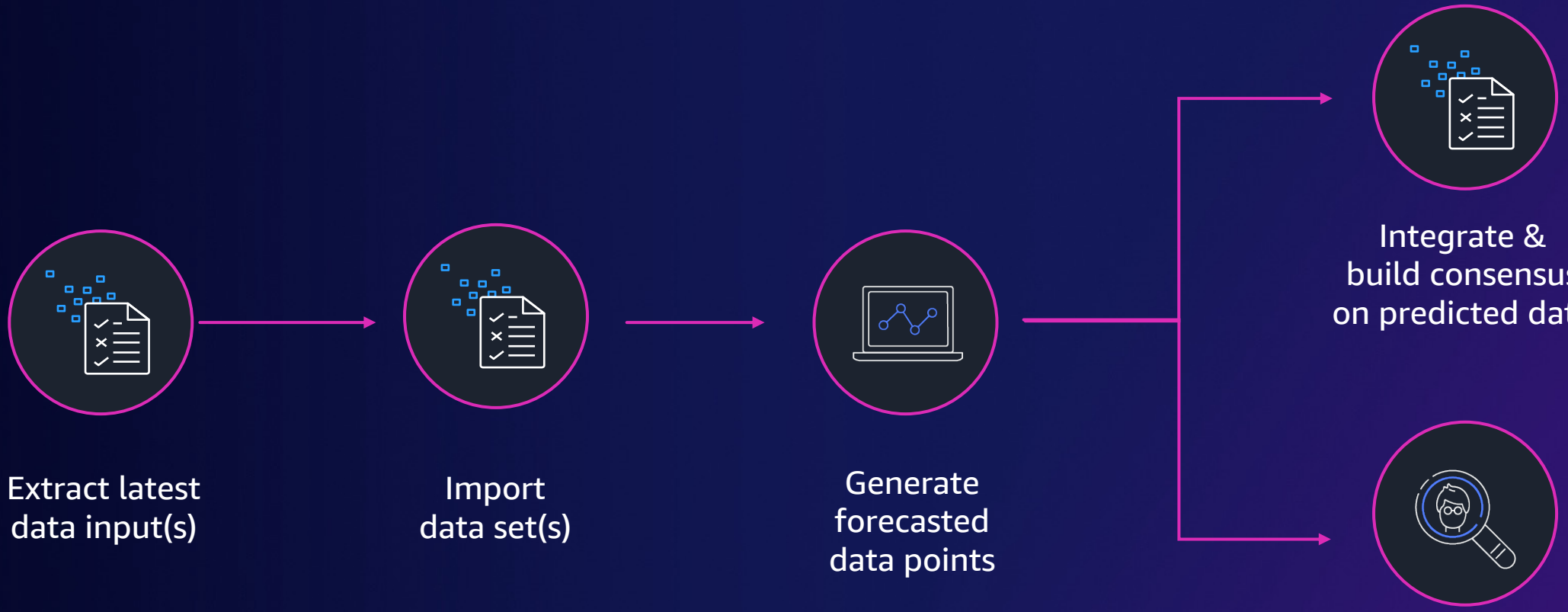
Explainability

- Quantify relative impact of each attribute
- Directional impact (+/–) of each attribute
- Model or time-series level explanations
- Supports drill down to specific time points for time series



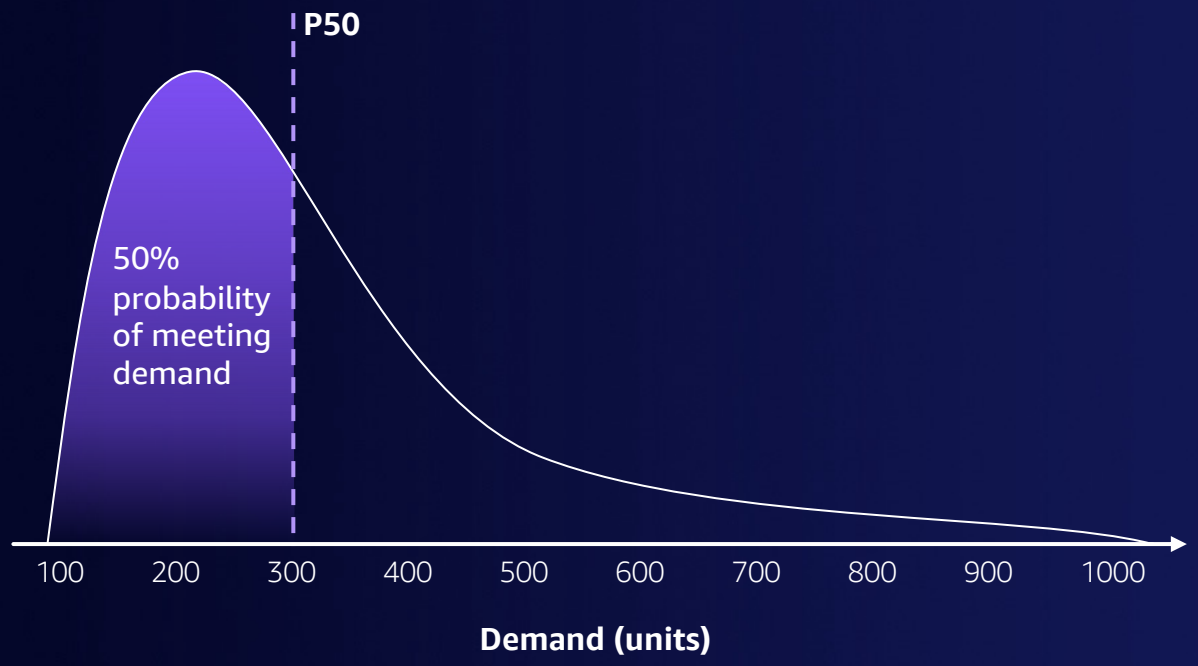
Forecasting workflow

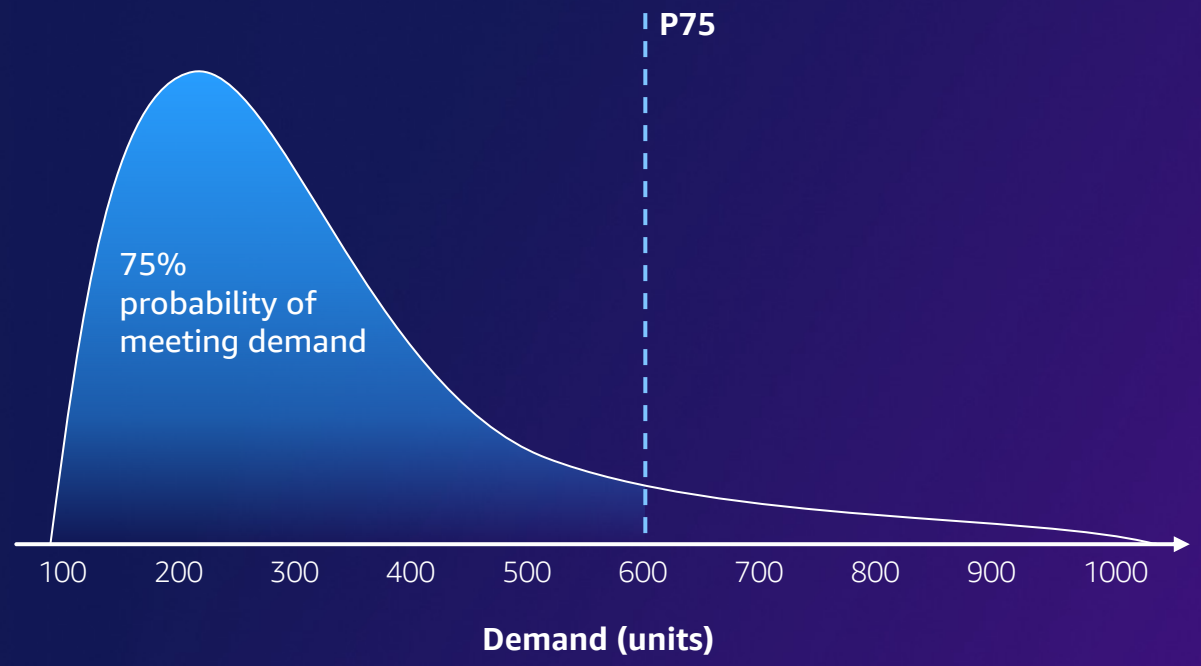
AUTOMATED FORECASTING BY SCHEDULES OR BY EVENTS



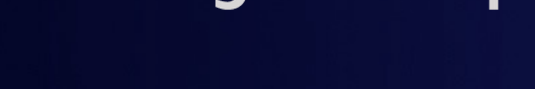
Balancing under-forecasting and over-forecasting

FORECAST AT VARIOUS QUANTILES OF THE PROBABILITY DISTRIBUTION





Forecasting customers



More Retail improved gross profits by 25%

TOP GROCERY RETAILER IN INDIA

Challenge

- Low shelf life.
- Large number of SKUs (1000+ items across 600+ stores).
- Wasted product directly impacts category profitability.

Solution

- Created 2 nationwide models for two different categories.
- Increased accuracy by learning across many time series.
- Out-performed ensemble of individual modes.
- Able to build rich features using related time series and item metadata to improve accuracy.
- Increased forecasting accuracy from 24% to 70% and reduced waste up to 30%.

aws

Foxconn saves $500K annually

MANUFACTURES SOME OF THE MOST WIDELY USED ELECTRONICS WORLDWIDE

Challenge

- Limited data science experience internally for Mexico factory.
- Having individual forecasts for each product is important to understand the mix of skills needed in workforce.

Solution

- Built a custom forecasting solution with Forecast in 2 months.
- Removed known anomalies.
- Used AutoML in Forecast to overcome limited background in time-series modeling.

FOXCONN

aws