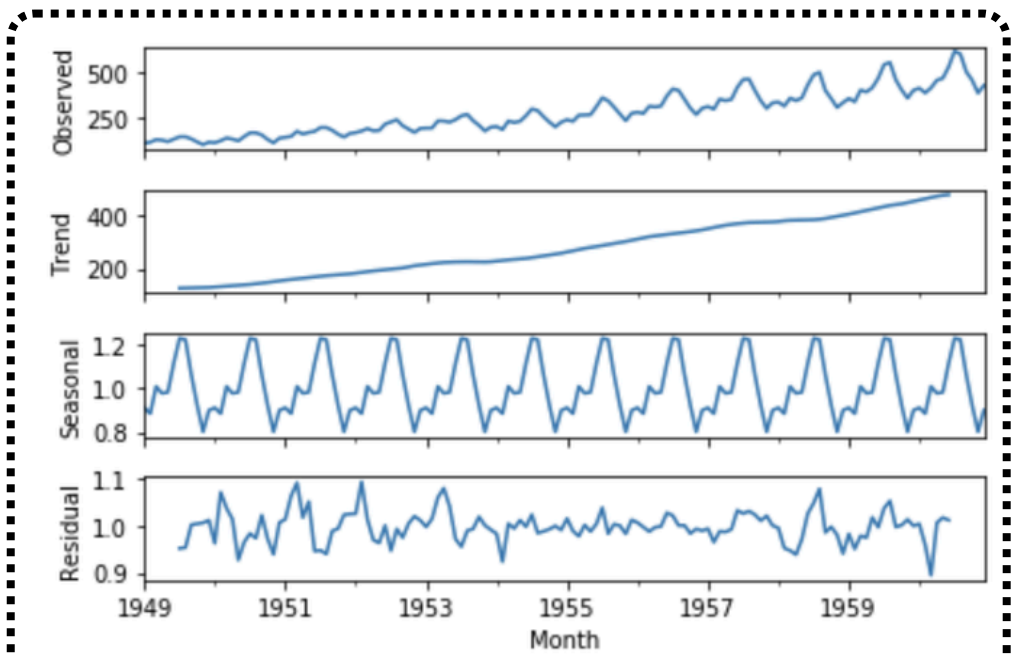
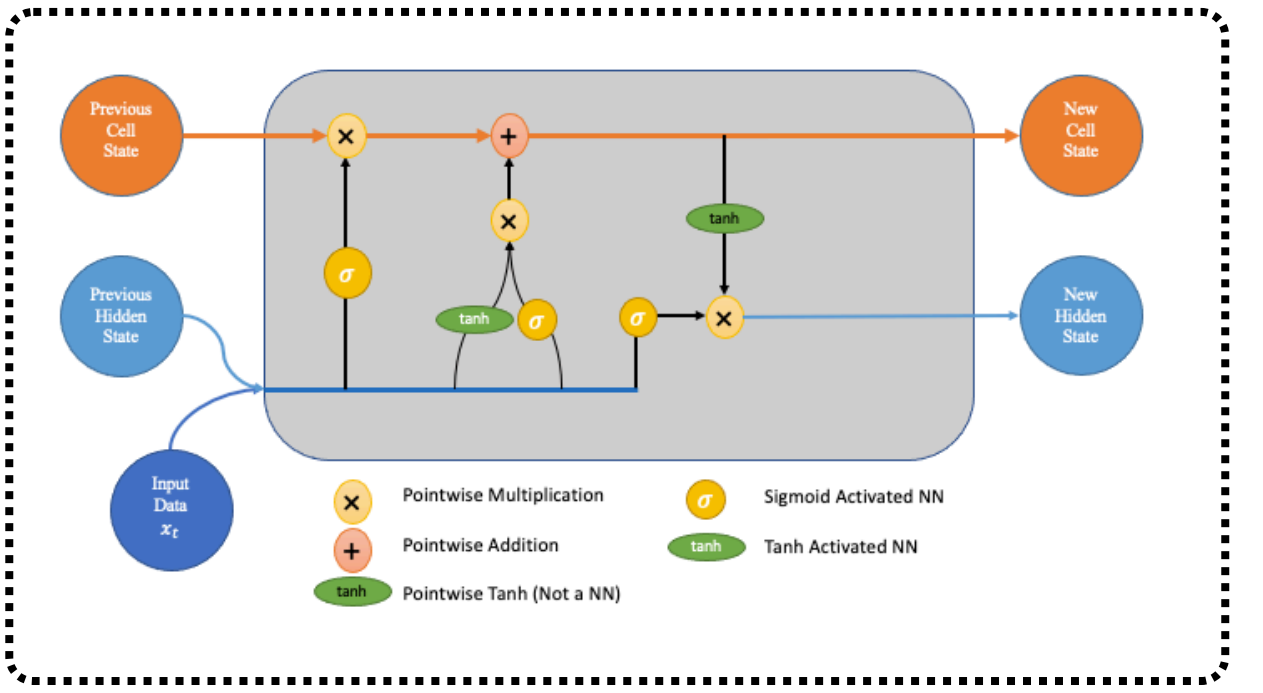
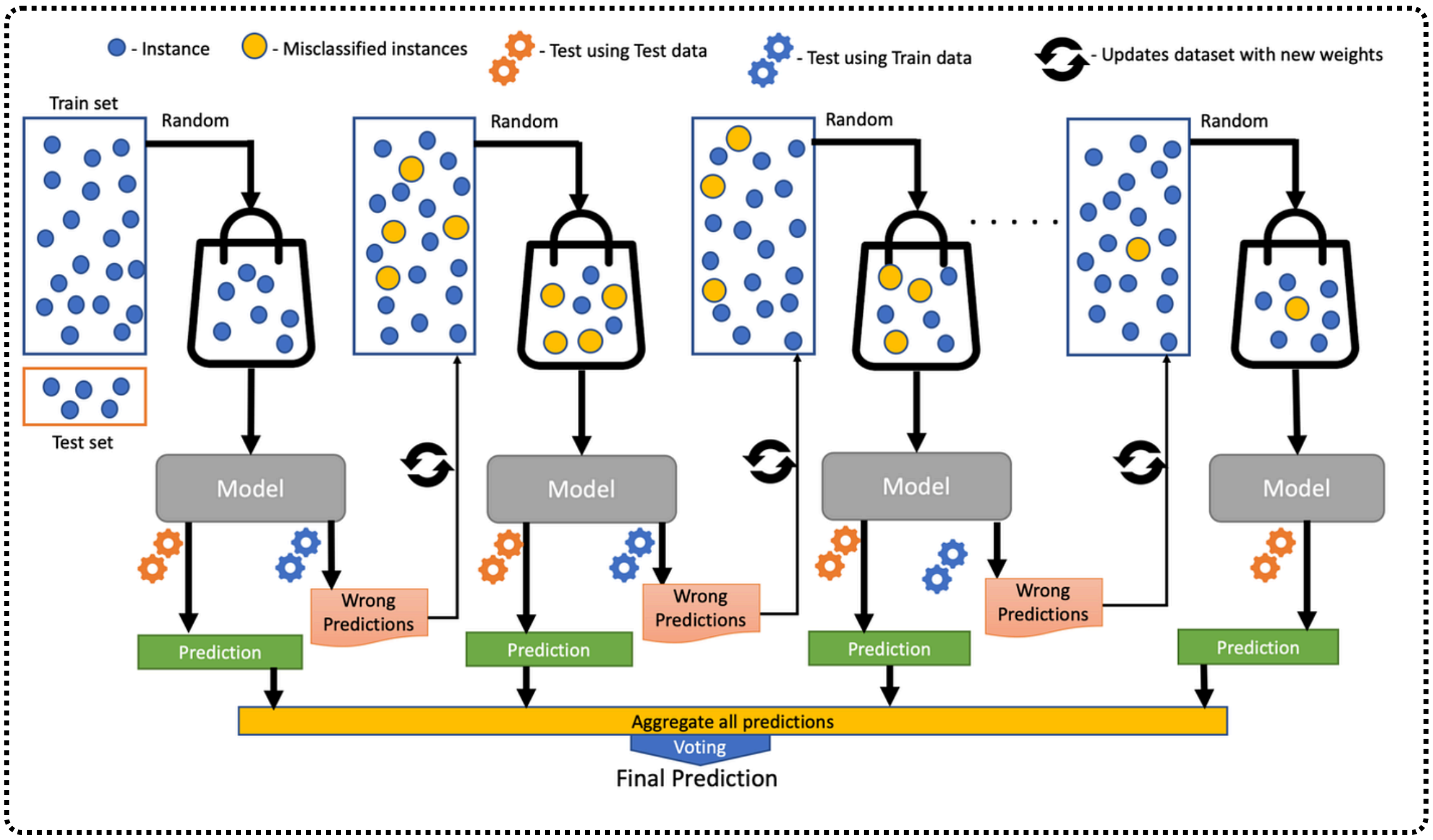
Analytics Vidhya

How IKEA Uses Data Science?







Demand Forecasting

Problem: Accurately forecast product demand at SKU × store × week level across global markets.

Data sources:

- Historical sales data
- Promotions and price changes
- Seasonality (month/week effects)
- Macroeconomic data (CPI, local income levels)
- Local events (back-to-school, Diwali, etc.)

Models:

- Classical: ARIMA, SARIMA for region-level trends
- Machine learning: XGBoost and LightGBM for SKU-level modeling
- Deep learning: LSTM and Temporal Fusion Transformers (TFT) for multivariate, multi-step forecasting

System architecture:

- Feature store (e.g. Feast) for time-indexed features
- Model deployment via MLflow + Airflow DAGs for retraining
- Forecasts written to Snowflake or BigQuery, served through BI dashboards (Tableau/PowerBI)

Use case: Drives production planning, warehouse stocking, and container load optimization.

Inventory Optimization

Objective: Minimize holding costs while maintaining high product availability.

Tech stack:

- Stochastic optimization using integer linear programming (ILP)
- Reinforcement learning (Q-learning) for adaptive inventory control policies in volatile regions
- Real-time simulations using SimPy or AnyLogic

Inputs:

- Forecasted demand
- Lead times
- Storage constraints
- Product substitutability

Output: SKU × location × reorder quantity

This feeds directly into ERP systems (SAP) and vendor order management tools.

Price Elasticity and Promotion Modeling

Goal: Estimate demand sensitivity to price changes.

Data:

- Historical pricing and unit sales
- Competitive pricing (web scraped or syndicated from providers)
- Promotion metadata (BOGO, % discount, duration)

Approach:

- Multivariate regression: Log-log model for elasticity estimation
- Bayesian hierarchical models: To pool data across similar products
- Uplift modeling: To assess promotion impact at the customer level

Use in production:

- Markdown planning
- Regional price optimization
- Dynamic pricing for online-only SKUs

Customer Behavior Modeling

Objective: Understand in-store and online purchase behavior to increase conversion and basket size.

Techniques:

- Market basket analysis using Apriori and FP-Growth for association rules
- Customer segmentation with K-Means, DBSCAN, and Gaussian Mixture Models
- Path analysis using Markov Chains to model customer journey through store or site
- Churn prediction with gradient boosting or ensemble models

Inputs:

- Transaction logs
- Loyalty card data
- Clickstream data (from IKEA.com and mobile apps)
- In-store sensor data (foot traffic, dwell time)

Deployment:

- Segments pushed into a CDP (Customer Data Platform)
- Real-time recommendations triggered via edge services

Product Recommendation Systems

Stack:

- Collaborative filtering (ALS, matrix factorization)
- Content-based filtering (product metadata: color, material, style)
- Hybrid models with embeddings via TensorFlow Recommenders
- Visual similarity via CNN-based image embeddings (ResNet or EfficientNet)

Personalization features:

- Geo-specific trending products
- Dynamic bundling (e.g. sofa + rug + lamp combo)
- Email/product listing personalization via Apache Beam pipelines

All trained and evaluated using precision@k, recall@k, and NDCG.

Geospatial Modeling for Store Expansion

Inputs:

- Population density
- Average income
- Competitor location
- Logistics cost matrix (distance to existing distribution centers)

Tools:

- QGIS for mapping
- PostGIS or BigQuery GIS for spatial queries
- XGBoost + SHAP for predictive modeling of store success probability

Used to simulate:

- Expected footfall
- Revenue ramp-up time
- Supply chain impact

Operational Infrastructure

Core stack:

- Data Lake: AWS S3 + Glue Catalog or GCP BigLake
- Orchestration: Apache Airflow for ETL, model retraining
- ML Platform: SageMaker or Vertex AI for training/deployment
- Analytics: Snowflake, BigQuery, dbt, Looker/Tableau

Versioning and reproducibility:

- Git + DVC for model and data versioning
- MLflow or Weights & Biases for experiment tracking

Privacy & compliance:

- GDPR compliance via user data anonymization pipelines
- Consent-based recommendation models (using privacy-first embeddings)