

COMPARATIVE ANALYSIS OF ARTIFICIAL INTELLIGENCE METHODS AND ALGORITHMS FOR INVENTORY DEMAND FORECASTING

Dilfuzakhon Abdullayeva
New Uzbekistan University

Abstract: Accurate demand forecasting is essential for inventory availability, working-capital efficiency and service continuity, yet retail demand is affected by promotions, seasonality, intermittent purchases and changing market conditions. This article presents a critical, narrative comparison of statistical, machine-learning and deep-learning approaches to inventory demand forecasting. ARIMA and intermittent-demand methods are assessed alongside Random Forest, XGBoost, LightGBM, long short-term memory networks, DeepAR and Temporal Fusion Transformers. The review examines modelling assumptions, information requirements, computational burden, interpretability, forecast uncertainty and relevance to replenishment decisions. Evidence from the M5 forecasting competition and methodological research indicates that no algorithm is uniformly superior: performance depends on the structure of the series, the availability of explanatory variables and the evaluation design. Tree-based ensembles offer a practical option for feature-rich retail data, while probabilistic neural models can support distribution-aware inventory decisions when sufficient related series are available. A conceptual hybrid forecasting framework and an evaluation protocol connecting forecast accuracy to inventory costs are proposed. The paper reports a literature-based synthesis rather than newly conducted experiments.

Keywords: deep learning; demand forecasting; gradient boosting; inventory management; machine learning; time series

INTRODUCTION

Inventory planning is a central operational problem in retail, manufacturing and distribution. A firm must hold sufficient units to meet uncertain demand while avoiding the financial and logistical consequences of excessive stock. Underestimating demand may lead to lost sales, emergency replenishment and deterioration in customer service. Overestimating it may increase storage requirements, tied-up working capital, obsolescence and product waste. Forecast quality is therefore important not as an isolated statistical objective, but as an input to decisions about ordering, replenishment and safety stock (Petropoulos et al., 2022).

Demand forecasting has become more difficult as product ranges expand and sales patterns become more heterogeneous. Daily sales are affected by promotional campaigns, price changes, calendar events, local preferences and supply limitations. Some items exhibit stable trends and regular seasonal cycles; others have frequent zero-sales periods and occasional demand spikes. A single forecasting system may consequently have to handle thousands of related but non-identical series, with different degrees of regularity and information availability (Makridakis et al., 2022a).

Classical forecasting models such as autoregressive integrated moving average (ARIMA) and exponential smoothing remain useful because they provide transparent, well-established statistical baselines. However, explanatory variables and nonlinear effects may be difficult to capture with a simple univariate specification. Machine-learning methods can model interactions among prices, promotions, lags and calendar attributes. Deep-learning models can exploit shared information across multiple series, while probabilistic approaches describe the uncertainty surrounding future

demand instead of producing only a single number (Hyndman & Athanasopoulos, 2021; Salinas et al., 2020).

These technical options create a selection problem. Greater architectural complexity does not automatically imply lower forecasting error; models differ in data needs, risk of overfitting, execution time, maintenance requirements and explainability. Moreover, a forecasting system with low average error may still cause costly stockouts if it fails to predict high-demand events or poorly represents the upper tail of the demand distribution. This distinction matters when forecast outputs are used for actual inventory control.

The objective of this paper is to compare the theoretical and practical capabilities of representative statistical, machine-learning and deep-learning methods for inventory demand forecasting. The contribution is a structured synthesis of method suitability, a transparent evaluation protocol and a conceptual framework connecting demand predictions with operational replenishment decisions. The article is a narrative review and does not claim original training runs, new benchmark scores or observed reductions in inventory costs.

METHODS

Review design and scope

The study uses a narrative and comparative literature-review approach. The analytical basis consists of foundational methodological publications for the selected algorithms, established research on forecast evaluation, and peer-reviewed studies concerned with retail demand and supply-chain forecasting. Priority is given to original method papers and documented forecasting competitions. This is not a systematic review: no comprehensive database search, predefined exclusion flow or pooled meta-analysis was undertaken. Therefore, the synthesis is intended to clarify modelling choices rather than estimate a universal performance ranking.

The methods were grouped into four categories: (1) classical statistical forecasting, including ARIMA and Croston-type intermittent-demand approaches; (2) tree-based machine learning, including Random Forest, XGBoost and LightGBM; (3) deep-learning approaches, including LSTM, DeepAR and Temporal Fusion Transformers (TFT); and (4) hybrid or ensemble combinations. Comparison criteria were predictive flexibility, capacity to use external regressors, treatment of sparse or intermittent demand, uncertainty representation, interpretability, computational requirements and suitability for large collections of product-level series.

Forecasting formulation and explanatory variables

Let $y(i,t)$ denote the number of units demanded for item i at time t . Given historical observations up to time t and a vector of available explanatory variables $x(i,t)$, the h -step point forecasting task can be represented by:

$$\hat{y}(i,t+h) = f_{\theta}[y(i,1:t), x(i,1:t), x_{\text{known}}(i,t+1:t+h)].$$

Here f_{θ} is a model with estimated parameters θ , and x_{known} refers only to predictors genuinely available at the forecast origin, such as a published holiday calendar or an already approved promotion schedule. Examples of historical features include lagged sales, moving averages, observed prices, stock availability and product characteristics. Future realized prices, promotions or sales must not enter training or testing as if they had been known in advance; doing so would create information leakage.

A crucial distinction is that recorded sales are not always equivalent to latent customer demand. During stockouts, observed sales may be censored by unavailable inventory. Treating every stockout-induced zero or low sale as low demand can bias the model downward. Consequently, stock-availability indicators and replenishment records should be integrated where possible, and missing-demand mechanisms should be considered explicitly.

Evaluation measures and validation design

For actual demand values $y_1, \dots, y_n$ and predictions $\hat{y}_1, \dots, \hat{y}_n$, two common point-forecast error measures are mean absolute error (MAE) and root mean squared error (RMSE):

$$\text{MAE} = (1/n) \sum |y_t - \hat{y}_t|; \text{RMSE} = \sqrt{[(1/n) \sum (y_t - \hat{y}_t)^2]}.$$

MAE describes average absolute deviation, whereas RMSE places greater weight on large errors. To compare series with different sales scales, mean absolute scaled error (MASE) is useful: forecast errors are divided by the in-sample mean absolute error of an appropriate naïve benchmark (Hyndman & Koehler, 2006). For hierarchical retail evaluation, the M5 competition employed the weighted root mean squared scaled error (WRMSSE), which accounts for aggregation levels and the relative importance of series (Makridakis et al., 2022b). No single metric alone describes all inventory consequences.

Data should be divided chronologically rather than through a simple random split. A rolling-origin design trains each method using observations available by a chosen date and tests predictions for subsequent periods. Several forecast origins should be used to evaluate changes across promotions, seasonal transitions and market disturbances. The same time windows, feature availability, forecasting horizons and preprocessing constraints must apply to every candidate model. Although particular cross-validation schemes can be valid under specific autoregressive assumptions, nonstationarity and future-information leakage require careful attention in practical retail forecasting (Bergmeir et al., 2018).

For models producing prediction distributions or quantiles, coverage and quantile-based loss should also be assessed. Operational evaluation should additionally measure stockout frequency, service-level attainment, inventory holding cost and replenishment cost under one explicitly specified inventory policy. In this paper, these metrics are proposed as a protocol; they were not computed from a new experimental dataset.

COMPARATIVE ANALYSIS OF FORECASTING ALGORITHMS

Statistical models: ARIMA and intermittent demand

ARIMA describes a time series through autoregression, differencing and moving-average components. With a suitable seasonal extension, it can capture repeated calendar structure and short-term temporal dependence. The method is attractive when sales histories are reasonably regular, observations are limited and model behaviour must remain readily interpretable. Exogenous terms can be incorporated through ARIMAX or related dynamic-regression formulations (Hyndman & Athanasopoulos, 2021).

Its disadvantages arise when relationships are highly nonlinear, structural breaks are frequent or numerous item-specific drivers must be fitted jointly. The model also requires suitable transformation, order selection and residual checks. For sparse spare parts and low-volume products, standard ARIMA may produce unstable or unhelpful estimates if long sequences of zeros dominate the training period.

Croston (1972) proposed an approach for intermittent demand that separately estimates nonzero demand sizes and the intervals between demand events. Such methods remain meaningful baselines for slow-moving inventories. They should not, however, be assumed automatically superior for every zero-inflated series; demand intermittency, trend, aggregation period and bias correction all influence their practical value.

Random Forest

Random Forest constructs an ensemble of decision trees using resampled observations and randomly selected predictor subsets. The resulting average prediction can capture nonlinear interactions without assuming a particular parametric relationship between explanatory variables

and demand (Breiman, 2001). In inventory forecasting, typical inputs include past sales, day-of-week indicators, promotion status, product category, store and item identifiers, price and recent rolling statistics.

The method is relatively straightforward to train and robust to many feature interactions. Feature importance and partial-dependence techniques can provide useful, though not causal, information about influential variables. A limitation is that time ordering is not represented automatically: appropriate lag and rolling-window construction is essential. Standard regression forests also have limited ability to extrapolate far beyond sales levels observed during training, which may be problematic during a new growth phase or an unprecedented promotion.

XGBoost and LightGBM

XGBoost is a scalable implementation of gradient-boosted decision trees. Trees are added sequentially to improve a specified loss function, with regularisation and computational optimisations designed to support accurate learning from structured features (Chen & Guestrin, 2016). It is especially relevant when price, promotion, seasonality, product and store attributes carry complementary predictive information. Proper tuning of tree depth, learning rate, sampling and regularisation is necessary to control overfitting.

LightGBM also uses gradient-boosted decision trees but introduces efficiency-oriented techniques, including gradient-based one-side sampling and exclusive feature bundling (Ke et al., 2017). It can be practical for high-dimensional, large-scale retail forecasting pipelines in which many item-location series are represented as one feature-rich training table. Its computational advantage in a particular deployment depends on data characteristics and implementation settings; speed should be measured rather than presumed.

Both gradient-boosting approaches can incorporate calendar, price and lag features and typically provide strong tabular-data baselines. Neither inherently produces a correctly calibrated future-demand distribution: probabilistic outputs require suitable loss functions, quantile methods, distributional models or calibration procedures. As with all supervised models, forecast-time feature leakage can create deceptively favourable validation results.

Long short-term memory networks

LSTM is a recurrent neural network architecture designed to learn temporal dependencies through a memory cell and gating operations that regulate information flow (Hochreiter & Schmidhuber, 1997). In demand forecasting it can represent relationships unfolding across many periods, particularly when sufficient historical sequences and relevant auxiliary variables are available. Its main advantage is the capacity to learn sequential representations with less manual specification of temporal relationships than some classical alternatives.

However, LSTM performance depends on training data, network design, optimisation settings and careful regularisation. A separate network for each sparse product series may be difficult to estimate reliably. Its computational demand can exceed that of classical methods, while the internal representation is generally less transparent than a simple regression model. Strong naïve and tree-based baselines are therefore needed before treating a neural forecast as an improvement.

DeepAR and Temporal Fusion Transformers

DeepAR trains an autoregressive recurrent model across many related time series and produces a predictive distribution conditional on observed history and available covariates (Salinas et al., 2020). Its global-learning design is relevant to large assortments: shared patterns can be learned across products rather than fitting every item in isolation. Distributional outputs are particularly valuable when order quantities must reflect asymmetric stockout and holding costs.

The quality of a DeepAR forecast still depends on the assumed probability family, series heterogeneity, scale handling and the amount of transferable information across items. For very short, unrelated or highly erratic series, pooling may offer limited benefits and calibration deserves special scrutiny. The presence of a prediction interval does not by itself guarantee reliable uncertainty estimates.

TFT combines recurrent processing, attention mechanisms, variable selection and gating to support interpretable multi-horizon forecasting with static and time-varying covariates (Lim et al., 2021). This structure is conceptually attractive when future calendar or promotion variables are available and several forecasting horizons are operationally important. The model can provide useful diagnostic indicators regarding feature relevance and temporal attention; these diagnostics should not be interpreted as definitive causal explanations.

TFT is comparatively complex to develop, tune and maintain. For smaller datasets or constrained computing environments, simpler statistical or gradient-boosting methods may offer a preferable accuracy-to-cost balance. The choice is fundamentally empirical and should be informed by the actual inventory decision context.

Table 1.

Qualitative comparison of forecasting approaches (author’s synthesis of the cited literature)

Method	Main strengths	Principal limitations	Suitable starting context
ARIMA / seasonal statistical models	Interpretable; established baselines; moderate data needs	Limited nonlinear flexibility without extensions	Regular, adequately observed demand
Croston-type methods	Designed for intermittent, nonzero demand events	Bias and structural-change sensitivity	Slow-moving or sporadic items
Random Forest	Nonlinear feature interactions; practical tabular baseline	Needs lag engineering; weak extrapolation	Mixed retail features; moderate data
XGBoost	Flexible regularised boosting; strong structured-data option	Tuning and leakage risks; uncertainty not automatic	Feature-rich item–store forecasting
LightGBM	Efficient boosting on many structured observations	Hyperparameter and sparsity sensitivities	Large retail training tables
LSTM	Learns sequential nonlinear patterns	Data- and training-intensive; opaque	Long, informative time series
DeepAR	Global learning; predictive distributions	Calibration and distributional assumptions	Many related demand series
TFT	Multi-horizon covariates and interpretable components	Substantial complexity and maintenance	Rich multivariate, multi-horizon tasks

Table 1 summarises architectural capabilities and practical trade-offs. It is not a table of measured accuracy scores, and its entries should not be interpreted as a universal performance

ranking. In particular, sparse demand, data leakage, stockout censoring and the availability of future explanatory variables can substantially change comparative outcomes.

RESULTS AND DISCUSSION

Evidence from forecasting research

The M5 Accuracy competition provides an important empirical reference because it used real retail sales organised into 42,840 hierarchical time series, including 30,490 bottom-level item–store series. Participants were evaluated using a common task and benchmark framework (Makridakis et al., 2022a, 2022b). The competition demonstrated that methods exploiting rich explanatory features and information sharing across related series can perform effectively on large retail forecasting problems. Nevertheless, the results reflect the specific dataset, horizon, metrics and rules of that competition; they do not prove that any single algorithm is best in all supply-chain environments.

The published M5 Accuracy analysis reported that the highest-ranked systems improved upon the strongest benchmark under the competition’s evaluation design (Makridakis et al., 2022b). This result supports considering machine-learning ensembles, but it does not replace direct evaluation of a company’s own demand data. The M5 Uncertainty competition separately evaluated probabilistic prediction, illustrating that distribution quality is a distinct objective from point-forecast accuracy (Makridakis et al., 2022c).

Earlier supply-chain research similarly identified opportunities for machine-learning techniques in settings with complex demand distortion and limited information sharing. Carboneau et al. (2008) examined neural-network, recurrent and support-vector methods in supply-chain demand forecasting. More broadly, the methodological synthesis by Petropoulos et al. (2022) emphasises selecting approaches appropriate to the available evidence, forecasting task and decision rather than relying on a single modelling family.

Matching algorithm families to demand characteristics

For relatively smooth, stable item-level demand with limited auxiliary data, seasonal statistical forecasting offers a defensible first baseline. When daily sales depend on prices, holiday effects and promotions, boosted trees should be tested because they can accommodate extensive engineered features and nonlinear interactions. For large product catalogues containing many related series, global neural approaches such as DeepAR may help by transferring statistical information across items. TFT warrants consideration when multi-horizon planning and mixed static, observed and known-future covariates justify its added complexity.

For intermittent demand, a naive model trained directly on zero-heavy sales can behave poorly even when its aggregate error seems acceptable. The modelling unit, forecast aggregation and stockout records become crucial. Where uncertainty about whether any demand will occur is substantial, two-stage approaches - first estimating event probability and then conditional positive demand - may also be worth testing alongside Croston-type and distributional alternatives. Such recommendations are hypotheses about suitability, not claims of measured superiority.

Interpretability is another selection criterion. Inventory managers must often explain why the system recommends larger orders during a particular period. ARIMA parameters, engineered predictors in boosted trees, and interpretable components in TFT can support different forms of explanation. However, a model interpretation shows association within a fitted system; it does not establish that changing a particular variable would causally alter demand.

Table 2.

Scenario-oriented model selection matrix (reasoned recommendations, not experimental rankings)

Observed forecasting context	Primary candidate methods	Required validation focus
Short, stable series; few explanatory variables	Seasonal naïve, ETS/ARIMA	Benchmark improvement; seasonal residuals
Sparse purchases and many zero periods	Croston variants; two-stage models	Demand occurrence; stockout bias; service cost
Rich calendars, prices and promotional variables	XGBoost, LightGBM, Random Forest	Feature availability at forecast origin; drift
Many related items with history across stores	Global neural methods; boosted trees	Cross-series transfer; cold start; calibration
Multi-horizon planning with known future events	TFT; multi-step boosting; DeepAR	Horizon-specific loss; interpretability; compute cost
High stockout penalties or asymmetric inventory cost	Calibrated quantile/distributional forecasts	Service level; quantile calibration; total cost

Linking forecasting performance with inventory decisions

Point forecasts answer the question of expected demand, whereas inventory management must also account for uncertainty, replenishment lead time and the relative costs of excess and insufficient inventory. In a simplified single-period newsvendor setting with underage cost C_u and overage cost C_o , the optimal order-up-to quantity corresponds to a conditional demand quantile at probability $C_u/(C_u + C_o)$, provided that the assumed cost and demand conditions hold. This makes calibrated predictive distributions operationally useful. For a multi-period replenishment system, order constraints, lead-time variability, review frequency, minimum order quantities and existing on-hand inventory must additionally be considered.

A forecasting algorithm can therefore be preferred under a combined criterion, conceptually written as:

$$J = \alpha \cdot L_{\text{forecast}} + \beta \cdot C_{\text{inventory}} + \gamma \cdot C_{\text{compute}},$$

where L_{forecast} denotes a normalised forecasting loss, $C_{\text{inventory}}$ denotes the inventory-policy cost produced by the forecast, C_{compute} denotes relevant computational or maintenance cost, and α , β and γ are decision-maker weights. These weights and terms must be defined on compatible scales prior to numerical optimisation. The equation is a conceptual multi-criteria objective proposed for subsequent experimental research, not an estimated model derived from this literature review.

A conceptual hybrid forecasting framework

The comparative findings suggest that combining different model families may be more defensible than imposing a universally preferred algorithm. A hybrid system could produce a statistical baseline and one or more machine-learning or neural forecasts, compare them through identical rolling-origin validation, and combine or select predictions using historical out-of-sample evidence. When suitable probabilistic methods are available, forecast uncertainty would then be passed to the replenishment policy along with point predictions. data validation precedes modelling because inventory records may contain stockout censoring, product substitutions, coding changes, returns and missing observations. Chronological splits then provide comparable evaluation windows for candidate forecasts. Selection is driven not only by point-error measures but also by forecast bias, uncertainty calibration, maintenance cost and the effect on replenishment decisions.

The system is completed by feedback monitoring: demand drift and deteriorating service outcomes can trigger model review and retraining.

One potential hybrid forecast takes a convex combination of two candidate forecasts, $\hat{y}_{\text{hybrid}} = w \cdot \hat{y}_{\text{stat}} + (1 - w) \cdot \hat{y}_{\text{AI}}$, with $0 \leq w \leq 1$. The weight would be estimated only from earlier validation periods and subsequently tested on untouched dates. More elaborate gating models could condition the weight on seasonality, sparsity and promotion regime, but their additional flexibility also increases the risk of overfitting. Accordingly, simple baselines and transparent validation are essential controls.

Limitations and implications for future research

This review is limited by its narrative selection of sources and lack of new experimental comparisons. The entries in the two tables express methodological properties and conditional model-selection judgements rather than effect sizes. Benchmark findings cannot be directly transferred to every national market, product category or enterprise resource-planning system. Data quality and operational constraints may dominate architectural differences between forecasting models.

A future empirical study should specify publicly reproducible data or appropriately anonymised enterprise records, item-level demand frequency, stockout flags, forecast origin dates and predefined evaluation horizons. It should compare seasonal naïve forecasts, ARIMA or intermittent-demand baselines, Random Forest, XGBoost, LightGBM, and selected global neural models using identical information sets. Repeated rolling-origin tests, uncertainty calibration and sensitivity analyses for product segments should be reported. Finally, an explicit replenishment simulation should assess service level, average on-hand stock, stockout events and total inventory cost. Only after this work can a particular hybrid system’s accuracy or economic advantage be asserted.

CONCLUSION

The comparative review indicates that inventory demand forecasting is not best approached as a contest in which model complexity alone determines quality. Statistical methods remain valuable for regular demand and transparent baselines; Croston-type methods address aspects of intermittent demand; tree-based ensembles provide flexible modelling of structured retail predictors; and deep-learning methods offer ways to learn temporal patterns and share information across many related series. DeepAR and TFT extend the analysis toward probabilistic and multi-horizon forecasting, respectively.

The most appropriate method depends on demand structure, data availability, the realism of forecast-time features, computational capacity and operational loss. Evidence from the M5 competitions reinforces the value of rigorous benchmarks and the separate evaluation of accuracy and uncertainty. A sound research design should preserve chronological separation, address censored sales caused by stockouts, and assess how forecasting outputs change replenishment performance rather than reporting a single error measure.

A conceptual hybrid framework has been outlined to integrate statistical baselines, artificial intelligence predictions, rolling-origin evaluation and inventory-cost criteria. Its effectiveness remains to be established experimentally. Future research should develop and test adaptive hybrid algorithms under changing demand, with transparent datasets, reproducible validation and a direct link between forecast uncertainty and inventory optimisation.

References

1. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140–1154. <https://doi.org/10.1016/j.ejor.2006.12.004>
4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
5. Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23(3), 289–303. <https://doi.org/10.1057/jors.1972.50>
6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
7. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
8. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
9. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30. <https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree>
10. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
11. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022a). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting*, 38(4), 1325–1336. <https://doi.org/10.1016/j.ijforecast.2021.07.007>
12. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022b). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. <https://doi.org/10.1016/j.ijforecast.2021.11.013>
13. Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., & Winkler, R. L. (2022c). The M5 uncertainty competition: Results, findings and conclusions. *International Journal of Forecasting*, 38(4), 1365–1385. <https://doi.org/10.1016/j.ijforecast.2021.10.009>
14. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., et al. (2022). *Forecasting: Theory and practice*. *International Journal of Forecasting*, 38(3), 705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
15. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>