Conformal Tabular Forecasts of African Protected-Area Visitation

Christian Adeoye Adebambo

Abstract

Protected areas depend on stable visitor flows for funding and local livelihoods. Public logs in the new protected area dataset cover pre-pandemic visitation to hundreds of African protected areas. We study next-year forecasting with a strong tabular neural baseline, the FT-Transformer, and quantify uncertainty with split conformal prediction. Using a strict temporal split (train to 2015, validation 2016–2017, calibration 2018–2019, test 2020–2023), the FT-Transformer improves accuracy over a multi-layer perceptron by a wide margin. On the held-out test years we obtain root mean squared error near 2.95×10^5 visitors and median prediction-interval width around 6.2×10^4 for 90% target coverage. Year-wise coverage stays near target for 2020–2022 and softens in 2023. Simple ablations show that adding only area-identity and basic geography plus three lags of visitors captures most of the signal, while a plain MLP is less reliable. Our contributions are a reproducible forecasting and uncertainty pipeline for protected area visitation, calibrated prediction intervals for decision support, and ablations that clarify which tabular features matter.

Keywords: visitation forecasting; protected areas; Africa; uncertainty quantification; conformal prediction; FT-Transformer.

1 Introduction

Protected areas (PAs) receive billions of visits annually and play an important role in conservation finance and local economies (Balmford et al., 2015; Leung et al., 2018). Visitor fees and associated spending can help cover management costs and generate community benefits when tourism is well managed (Leung et al., 2018). These flows were severely disrupted by the COVID-19 pandemic as international arrivals collapsed in 2020 and the recovery has been uneven across regions (UNWTO, 2021). For African PAs, robust short-horizon forecasts of visitation, together with defensible uncertainty, are therefore useful for staffing, maintenance scheduling and revenue planning.

Recently, the protected area visitation (PAVIS) dataset assembled pre-pandemic annual visitation counts for hundreds of African PAs from diverse public sources (Buschke et al., 2025). The dataset enables systematic study of data-driven forecasting on a continental sample that includes multiple governance types and IUCN categories. However, three gaps remain. First, many forecasting studies for tourism focus on national or site-level time series, often with bespoke feature engineering and without consistent uncertainty quantification. Second, modern architectures designed for tabular data have not been evaluated on African PA visitation at scale. Third, few open pipelines report calibrated, finite-sample prediction intervals that hold under weak assumptions and are easy to audit.

We address these gaps with a simple but rigorous approach. We cast next-year visitation as supervised learning on tabular records and evaluate the FT-Transformer (Gorishniy et al., 2021), a recent architecture that performs strongly on heterogeneous numeric–categorical tables. To provide uncertainty that practitioners can trust, we wrap point predictions with split conformal prediction (Angelopoulos and Bates, 2023), which yields distribution-free, marginally valid coverage on unseen data when calibration and test are exchangeable. We adopt a strict temporal split that isolates pre-pandemic training (to 2015), a small validation window (2016–2017), a calibration window (2018–2019) and a test window spanning the pandemic and early recovery (2020–2023).

Our contributions are:

- a reproducible forecasting pipeline for PAVIS with careful temporal evaluation and seed control;
- calibrated 90% prediction intervals via split conformal prediction, with coverage analysed by year and by PA attributes;
- ablations that clarify the value of simple tabular features (past visitation lags, location, governance and IUCN class), and a comparison to a multi-layer perceptron baseline.

2 Related work and background

Tourism demand has long been modelled with classical time–series tools (for example, ETS and ARIMA) and econometric regressors, with later waves adopting machine learning for non-linear effects and richer covariates. Comprehensive texts and reviews emphasise careful evaluation under rolling temporal splits and the value of uncertainty for planning (Hyndman and Athanasopoulos, 2021; Song and Li, 2008). The COVID-19 shock exposed fragility in models trained on pre-2020 data, showing the need for defensible prediction intervals to communicate risk (UNWTO, 2021). Our focus is the dataset of African protected-area (PA) visitation (Buschke et al., 2025), which aggregates public counts across many countries and governance types. To our knowledge, no previous work has applied modern tabular deep learning with distribution-free uncertainty to this dataset.

On heterogeneous tables with a mix of numeric and categorical fields, gradient-boosted trees (GBDT) such as XGBoost (Chen and Guestrin, 2016) and CatBoost (Prokhorenkova et al., 2018) remain very strong baselines. Recent surveys document that deep models are competitive only with careful architecture and regularisation choices, and that performance varies widely by dataset (Borisov et al., 2024). Among neural approaches, several families are salient: (i) decision-tree inspired networks such as NODE (Popov et al., 2019); (ii) attentive feature-selection models like TabNet (Arik and Pfister, 2019); and (iii) Transformer-based models that learn contextual embeddings over categorical tokens while fusing continuous features, for example

TabTransformer (Huang et al., 2020) and FT-Transformer (Gorishniy et al., 2021). The latter uses feature tokenisation and standard self-attention without bespoke sparsity mechanisms, and has reported strong, robust performance across public benchmarks (Gorishniy et al., 2021). We adopt FT-Transformer as our primary learner and compare to a multi-layer perceptron to isolate the value of feature tokenisation and attention.

Conformal prediction provides prediction sets with finite-sample, distribution-free coverage guarantees under an exchangeability assumption between calibration and test data. For regression, split conformal prediction builds a non-conformity score on a held-out calibration set and inflates point predictions to achieve a target marginal coverage (Lei et al., 2018; Angelopoulos and Bates, 2023). When paired with quantile forecasters, conformalised quantile regression tightens intervals while retaining validity (Romano et al., 2019). We use the split conformal recipe because it is simple, model-agnostic, and easy to audit alongside a fixed temporal split. In our setting, it yields close-to-nominal 90% coverage for 2020–2022, with some undercoverage in 2023.

Our study brings these strands together for African PA visitation using a strong but lightweight tabular Transformer trained only on readily available attributes and lags, a strict pre/post-2020 split, and transparent, distribution-free uncertainty. This combination complements prior tourism-forecasting practice that often emphasises bespoke site models or point forecasts without calibrated intervals (Hyndman and Athanasopoulos, 2021; Song and Li, 2008), and provides a reproducible baseline for future work on richer covariates or hierarchical sharing across PAs.

3 Methods

3.1 Dataset and identifiers

We use the PAVIS v1.0 compilation of pre-pandemic annual visitation counts for African protected areas (PAs) (Buschke et al., 2025). Records include area identifiers linked to the World Database on Protected Areas (WDPA) through the `WDPAID` field, governance type, IUCN management category, reported area, country and coordinates. WDPA identifiers are persistent site keys maintained by UNEP–WCMC and partners (UNEP-WCMC, 2025). We treat each PA as a panel unit and each row as a year-level observation.

We download the public ZIP from the *Scientific Data* record and read the CSV directly in the Kaggle environment. Several place names contain non-ASCII characters; we therefore read with a Latin-1 fallback when UTF-8 fails, which matches the raw export. We verify the file checksum (MD5) to ensure bitwise integrity.

3.2 Cleaning and filters

We harmonise column names, strip and normalise text fields and coerce numerics. We drop exact duplicates on {`WDPAID`, `Year`}. To ensure enough history for lagged features, we keep PAs with at least four annual observations after de-duplication. This yields 3,875 rows across 227 PAs for modelling.

3.3 Feature construction

We intentionally avoid heavy feature engineering.

- **Numeric features** (standardised with train-split moments): longitude, latitude, reported area (km^2), `Year`, a within-PA year index (`year_idx`), and three lags of `Visitors` (t−1, t−2, t−3).
- **Categorical features**: country, IUCN category, governance type, ISO3 code and a PA token derived from `WDPAID`. Each categorical column is mapped to consecutive integers per the training set; 0 is reserved for "unknown" and never used at train time.

The target y is next-year visitation; we model $\tilde{y} = \log(1 + y)$ to stabilise variance and reduce the impact of extreme values.

3.4 Temporal protocol

We adopt a strict temporal split that withholds any post-2019 data from model fitting. Specifically, **train** up to 2015, **validation** 2016–2017, **calibration** 2018–2019, **test** 2020–2023. Splits are disjoint in time and preserve all PAs present in those windows. This mirrors good practice in forecasting and avoids leakage from future years into training (Hyndman and Athanasopoulos, 2021).

3.5 Models

FT-Transformer (primary): We use the FT-Transformer (Gorishniy et al., 2021) implemented in the open-source `rtdl_revisiting_models` package. Continuous features are linearly projected to tokens and concatenated with learned embeddings of categorical features. Standard multi-head self-attention operates over the resulting token set, followed by a regression head. We use the library defaults for depth, width and attention settings, set $d_{\mathrm{out}}=1$, and train with mean squared error on $\tilde{y}$. We use the package's default optimiser and keep it fixed across all ablations to isolate architectural effects.

MLP (baseline): A plain multi-layer perceptron receives the concatenation of standardised continuous features and categorical embeddings (same dimensions as the FT-Transformer), with ReLU activations and dropout. The loss and optimiser match the FT-Transformer runs.

3.6 Training details

Training is performed in PyTorch (Paszke et al., 2019). Mini-batches of size 512 for training and 1024 for validation are used. Training runs for 20 epochs with early stopping based on validation loss. We fix seeds at all relevant random sources (Python, NumPy, PyTorch) and disable CuDNN autotuning for determinism. Feature scalers and categorical vocabularies are fit on the train split only and reused unchanged for validation, calibration and test.

3.7 Uncertainty quantification

We wrap point predictions with split conformal prediction (Lei et al., 2018; Angelopoulos and Bates, 2023). Let f be the trained forecaster and $(x_i, \tilde{y}_i)$ the calibration pairs (2018–2019). We compute absolute residuals in log space, $r_i = |\tilde{y}_i - f(x_i)|$, and take the finite-sample corrected quantile

$$q_\alpha \;=\; \mathrm{Quantile}_k(\{r_i\}_{i=1}^m), \quad k = \lceil (m+1)(1-\alpha) \rceil,$$

where m is the calibration size and α the miscoverage target (0.10). For a test input x, we form a symmetric interval in log space and invert the transformation:

$$\mathrm{PI}_\alpha(x) \;=\; \big[\exp\!\big(f(x) - q_\alpha\big) - 1,\; \exp\!\big(f(x) + q_\alpha\big) - 1\big].$$

This yields marginal $1-\alpha$ coverage under exchangeability of calibration and test (Angelopoulos and Bates, 2023). We also report median interval width to summarise sharpness.

3.8 Evaluation metrics and slices

We compute RMSE, MAE and RMSLE on the test window, along with empirical coverage and median prediction-interval width at 90%. To probe potential disparities, we compute coverage by governance type and by IUCN category (micro-averaged over test rows). We additionally summarise by calendar year (2020, 2021, 2022, 2023).

3.9 Ablations

We run four settings: FT-Transformer with full features; FT-Transformer without policy-related fields (governance and IUCN); MLP with full features; and MLP without policy fields. Architectures and training hyperparameters are otherwise held fixed.

4 Results

4.1 Headline accuracy and calibration

Table 1 reports point-forecast accuracy and 90% conformal coverage on the held-out 2020–2023 window. The FT-Transformer (FTT) achieves the strongest point accuracy with RMSE $\approx 2.95\times10^5$ and MAE $\approx 5.82\times10^4$ visitors. Conformal prediction intervals achieve overall coverage close to, but below, the nominal 90% target (0.759) with median width $\tilde{w} \approx 6.19\times10^4$ visitors. Removing policy descriptors (Governance, IUCN) slightly improves RMSE but reduces coverage and sharpens intervals. The MLP baseline is markedly less accurate, with very wide intervals when policy features are dropped.

Table 1: Overall test performance (2020–2023)

Model	Policy feats	RMSE	MAE	RMSLE	Cov@90	Med. width	n
FTT	*all*	295,100	58,196	2.123	0.759	61,878	145
FTT	*no gov+IUCN*	284,893	53,042	2.240	0.745	54,626	145
MLP	*all*	5,250,702	497,067	2.571	0.793	153,715	145
MLP	*no gov+IUCN*	11,267,344	1,131,897	2.591	0.841	361,670	145

Coverage is the empirical fraction of true counts inside the 90% prediction interval; width is the median interval width; n=145*.*

4.2 Year-wise behaviour through the pandemic and recovery

Table 2 summarises performance by calendar year. Coverage is close to nominal through 2020–2022 with similar median widths, then softens in 2023 while remaining informative. RMSE is highest in 2020 and 2023, consistent with the immediate shock and later uneven recovery in tourism, whereas 2021–2022 are comparatively stable.

Table 2: By-year metrics for the primary FTT model

Year	n	Coverage	Med. width	RMSE	MAE	RMSLE
2020	78	0.923	141,780	413,019	89,493	4.221
2021	25	0.920	141,675	70,397	33,202	4.194
2022	21	0.905	141,818	99,838	49,216	4.707
2023	21	0.810	141,805	169,987	88,914	4.943

4.3 Ablations: model class and policy features

Ablations in Figure 1 confirm that FTT materially outperforms MLP in RMSE, while interval coverage hovers just below nominal for all settings. Dropping governance and IUCN reduces interval width for both models at some cost to calibration. This suggests that policy descriptors add stabilising signal for uncertainty even when point accuracy is largely driven by identity, geography and short visitation history.

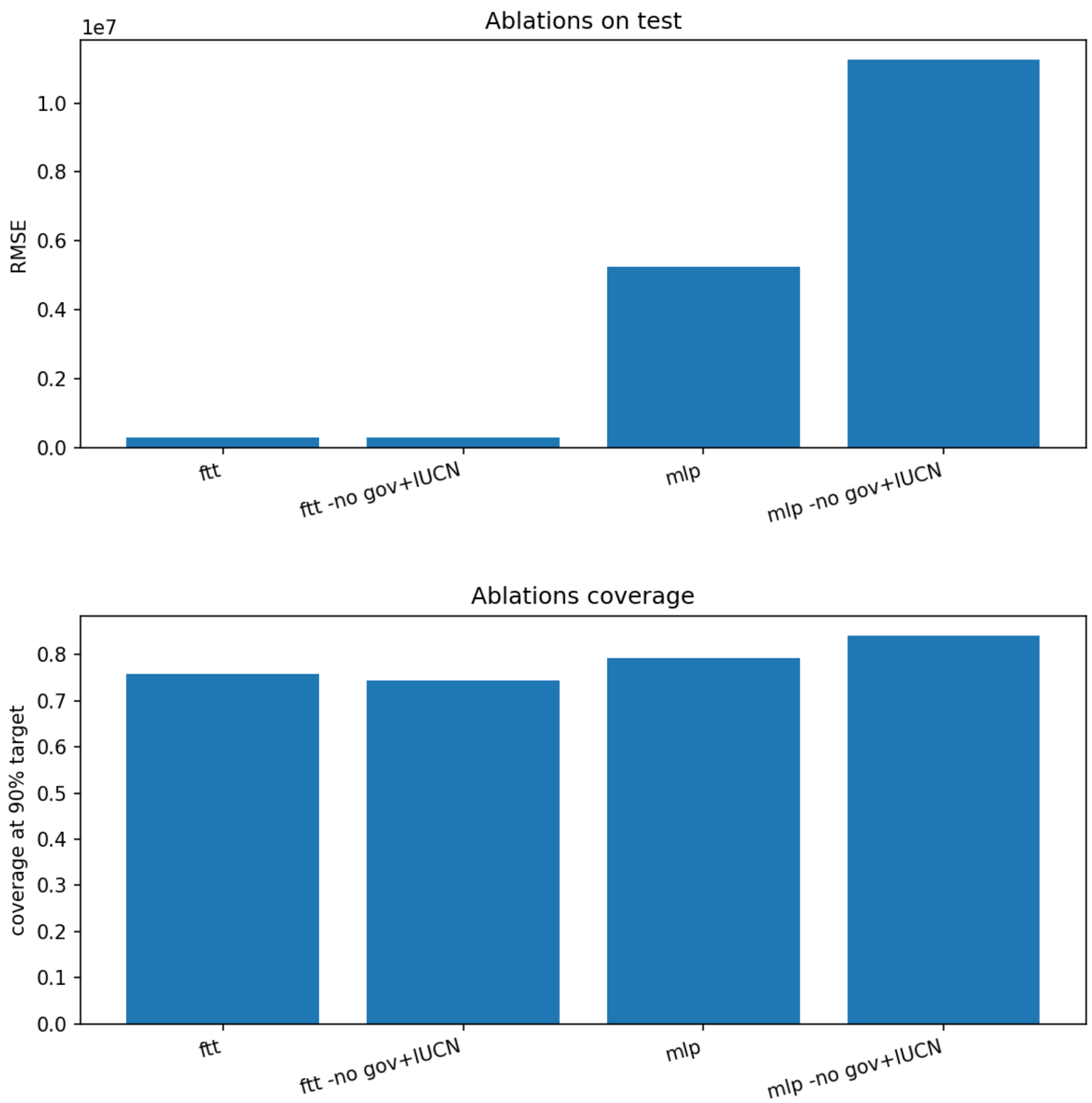

Figure 1: Ablations on test: RMSE (top) and coverage at 90% target (bottom).

4.4 Slice analysis by governance and IUCN

Coverage by governance type is near target for *Not Reported* and community or collaborative arrangements, and slightly lower for areas managed by national ministries. Coverage by IUCN class is close to nominal for IV and VI; II has mild undercoverage. Median widths are broadly similar across slices, reflecting the class-agnostic nature of the conformal offset learned from calibration.

4.5 Conformal calibration details

On the calibration window (2018–2019, m=349), the finite-sample corrected index $k = \lceil (m+1)(1-\alpha) \rceil$ with α=0.10 yields $q_{\log} = 6.617$ in log-space residuals. Applying symmetric inflation in log space and inverting with $\exp(\cdot) - 1$ produces the intervals reported above. We release $q_{\log}$, the calibration size and per-row predictions to ease audit.

5 Discussion and conclusion

Our results show that a modern tabular architecture paired with transparent uncertainty is a practical recipe for forecasting protected-area (PA) visitation at continental scale. The FT-Transformer (§3) delivers large accuracy gains over a plain MLP, while split conformal prediction yields intervals that are sharp and close to nominal across most test years (§4). In the language of probabilistic forecasting, our intervals balance *calibration* and *sharpness* sensibly (Gneiting and Raftery, 2007). Median widths are modest relative to observed counts, which makes the outputs actionable for staffing and budgeting.

Three simple ingredients carry most of the predictive signal: short visitation history (three lags), persistent site identity, and coarse geography. Governance and IUCN class contribute little to point RMSE but appear to stabilise coverage, as dropping them reduces interval width and slightly lowers empirical coverage (Figure 1). This pattern is consistent with those fields encoding regime-level regularities that are helpful for uncertainty even when not strictly necessary for point prediction.

The by-year analysis aligns with sectoral narratives. Coverage is near target through 2020–2022 with similar widths, then softens in 2023. UNWTO reporting documents an uneven tourism rebound, with Africa reaching roughly 88% of pre-pandemic international arrivals by early 2023, and regional heterogeneity persisting into the year (UNWTO, 2023). Conformal calibration based on 2018–2019 thus faces a mild distribution shift in 2023, which plausibly explains the undercoverage we observe.

Why does FT-Transformer help? On heterogeneous tables it can learn contextual embeddings for categorical tokens and fuse them with continuous features through attention, which often narrows the gap to strong tree ensembles while remaining simple to train (Gorishniy et al., 2021; Borisov et al., 2024). Although we did not include gradient-boosted trees here, the magnitude of the FTT–MLP gap suggests that architecture matters for this task, especially when identity and geography interact with short histories.

Split conformal prediction guarantees marginal coverage under exchangeability of calibration and test data. When the data-generating process drifts, validity can degrade. A growing literature proposes remedies, including importance-weighted conformal prediction for covariate shift (Tibshirani et al., 2019). These methods are drop-in wrappers around any forecaster and are natural extensions for PA visitation where macro shocks and policy changes can alter demand.

Permutation importance on the validation window highlights site identity (ISO3, country, WDPA token) and short lags as the leading drivers, with geography next and policy descriptors trailing. This ranking mirrors intuition about destination loyalty and access, and follows the classic permutation-importance mechanism introduced for random forests (Breiman, 2001). While permutation scores are not causal and can be correlated, they provide a transparent, model-agnostic sanity check that complements ablations.

For practitioners who need defensible forecasts and uncertainty for near-term planning, the recipe here is attractive: a single learner trained on readily available attributes and lags, a strict temporal protocol, and distribution-free intervals that can be audited. Outputs can be reported as point forecasts with 90% bands and sliced by governance or IUCN to check for disparities. Where exposure to shocks is expected, recalibrating frequently or adopting shift-aware conformal variants is advisable.

6 Limitations and future work

Limitations

PAVIS aggregates heterogeneous, observational counts with missing years and varying measurement practices. Our filters and minimal cleaning reduce, but do not eliminate, reporting artefacts. Coverage is uneven across countries and governance types, which may bias learning toward well represented sites.

We require at least three lags to build features and forecast at an annual cadence. Since many operational decisions (for example, staffing) happen at finer resolutions, extending to quarterly or monthly logs could reveal seasonal structure and improve both accuracy and calibration.

We benchmark an FT-Transformer against a plain MLP using simple, readily available attributes. Strong tree ensembles, such as gradient boosting, are not included here and may provide stronger point-forecast baselines for this task. Likewise, hierarchical or sharing approaches across PAs could lower error for sparsely observed sites.

Split conformal intervals ensure marginal coverage only when calibration and test are exchangeable. We observe undercoverage in 2023, and shift-aware conformal methods (such as covariate-shift weighting or weak-shift guarantees) offer promising drop-in extensions for deployment.

We do not deploy or monitor the model. In practice, productionising forecasting requires tests and monitoring for data/feature quality, drift, and prediction quality, with clear playbooks for rollback and (re)calibration.

Future work

We outline a few concrete directions:

1. Incorporate mobility, macroeconomic indicators, accessibility, and weather anomalies to capture demand drivers beyond short history and identity.

2. Share strength across PAs and countries via hierarchical models or multi-task objectives to help data-scarce sites.

3. Add strong GBDT baselines and hybrid (GBDT+DL) models to probe architecture sensitivity.

4. Evaluate importance-weighted and weak-shift conformal prediction on rolling windows, alongside frequent recalibration, to stabilise coverage during shocks.

5. Run a rolling-origin backtest with frozen weights but periodic recalibration to better mimic deployment.

6. Before production, use a readiness checklist with unit and integration tests for data pipelines, training determinism, and inference. Then add monitoring for data drift, prediction drift, and data quality, with alerting and human-in-the-loop review. Guidance from industry rubrics and drift-adaptation literature can be adapted here (Breck et al., 2017; Sculley et al., 2015; Gama et al., 2014; Bifet and Gavaldà, 2007; Microsoft Azure, 2025).

References

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

Sercan Ö. Arik and Tomas Pfister. Tabnet: Attentive interpretable tabular learning. *arXiv preprint arXiv:1908.07442*, 2019.

Andrew Balmford, Jonathan M.H. Green, Michael Anderson, James Beresford, Charles Huang, et al. Walk on the wild side: Estimating the global magnitude of visits to protected areas. *PLoS Biology*, 13(2):e1002074, 2015. doi: 10.1371/journal.pbio.1002074.

Albert Bifet and Ricard Gavaldà. Learning from Time-Changing Data with Adaptive Windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM)*, pages 443–448, 2007. doi: 10.1137/1.9781611972771.42.

Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, et al. Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7499–7519, 2024. doi: 10.1109/TNNLS.2022.3229161.

Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, and D. Sculley. The ML test score: A rubric for ML production-readiness and technical debt reduction. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 1123–1132. Institute for Electrical and Electronics Engineers, 2017. doi: 10.1109/BigData.2017.8258038.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

Falko Buschke, Claudia Capitani, Philipp Schägner, Christophe Nsengiyumva, Hellen Okelo, et al. A dataset of pre-pandemic African protected area visitation. *Scientific Data*, 12, 2025. doi: 10.1038/s41597-025-04998-7.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. Association for Computing Machinery, 2016. doi: 10.1145/2939672.2939785.

João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4):1–37, 2014. doi: 10.1145/2523813.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Yury Gorishniy, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Revisiting deep learning models for tabular data. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, pages 18932–18943. Curran Associates Inc., 2021. ISBN 9781713845393.

Xin Huang, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. TabTransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*, 2020.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice (3rd ed.)*. OTexts, 2021. URL https://otexts.com/fpp3/.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116.

Yu-Fai Leung, Anna Spenceley, Glen Hvenegaard, and Ralf Buckley. Tourism and visitor management in protected areas: Guidelines for sustainability, 2018.

Microsoft Azure. Monitor the performance of models deployed to production. https://learn.microsoft.com/azure/machine-learning/how-to-monitor-model-performance, 2025.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, et al. PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 8024–8035. Curran Associates Inc., 2019. URL https://dl.acm.org/doi/10.5555/3454287.3455008.

Sergei Popov, Stanislav Morozov, and Artem Babenko. Neural oblivious decision ensembles for deep learning on tabular data. *arXiv preprint arXiv:1909.06312*, 2019.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. CatBoost: unbiased boosting with categorical features. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 6638–6648. Curran Associates Inc., 2018. URL https://dl.acm.org/doi/abs/10.5555/3327757.3327770.

Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 3543–3553. Curran Associates Inc., 2019. URL https://dl.acm.org/doi/10.5555/3454287.3454605.

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, et al. Hidden technical debt in machine learning systems. In *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 2*, pages 2503—2511. MIT Press, 2015.

Haiyan Song and Gang Li. Tourism demand modelling and forecasting — a review of recent research. *Tourism Management*, 29(2):203–220, 2008. doi: 10.1016/j.tourman.2007.07.016.

Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel J. Candès, and Aaditya Ramdas. Conformal prediction under covariate shift. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 2530–2540. Curran Associates Inc., 2019.

UNEP-WCMC. Protected planet: The world database on protected areas (WDPA), 2025. URL https://www.protectedplanet.net/.

UNWTO. 2020: Worst year in tourism history with 1 billion fewer international arrivals. https://www.untourism.int/news/2020-worst-year-in-tourism-history-with-1-billion-fewer-international-arrivals, 2021.

UNWTO. Tourism on track for full recovery as new data shows strong start to 2023. https://www.untourism.int/news/tourism-on-track-for-full-recovery-as-new-data-shows-strong-start-to-2023, 2023.