Analyst-Driven XAI for Time Series Forecasting: Analytics for Telecoms Maintenance

JAMES BARRETT, University of the West of England, United Kingdom
PHIL LEGG, University of the West of England, United Kingdom
JIM SMITH, University of the West of England, United Kingdom
CHIP BOYLE, Ribbon Communications, USA

Time series forecasting facilitates real-time anomaly detection in telecom networks, predicting events that disrupt security and service. Current research efforts have been found to focus on new forecasting libraries, more rigorous data cleaning methods, and model hyperparameter tuning, although we believe human in-the-loop system approaches are not well applied to the domain of time series forecasting. We explore the usage of a model investigation tool to enable an interactive machine learning process that allows the interrogation of modern forecasting models to enable effective model management techniques. This research aims to demonstrate the usage of an interactive forecasting ensemble tool that enables a user to interrogate time series data, uncover insights in data predictions to make choices, and adjust a model accordingly. Through comparative testing and an analysis of existing model management strategies, we propose that enabling greater levels human-machine teaming via our tool promotes the ability to "catch" mistakes and oversights based on the assumptions of existing time series forecasting methods.

Additional Key Words and Phrases: Time Series Forecasting, Anomaly Detection, Human-in-the-Loop Machine Learning, Explainable Artificial Intelligence, Telecoms Data Analysis.

ACM Reference Format:
James Barrett, Phil Legg, Jim Smith, and Chip Boyle. 2024. Analyst-Driven XAI for Time Series Forecasting: Analytics for Telecoms Maintenance. 1, 1 (February 2024), 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Explainability of Machine Learning (ML) models, including multi-variate time-series forecasting methods, is fundamental for well-informed decision-making [9] [11] to ensure that models are performing as expected based on the available data. Complex systems, including telecommunication networks, require the understanding of multi-variate time series data to forecast and mitigate security and service issues, ranging from billing fraud [2] to network performance [1]. However, telecommunication networks are susceptible to external factors that may not be characterised within existing data attributes of a learnt model, including the behaviours of human users and the occurrence of external real-world events. Explainable Artificial Intelligence (XAI) is a human-machine collaborative process [5], bringing together machine learning models and visualisation techniques, to provide greater reasoning for a human analyst on the underlying properties of *why* a model has made such a prediction and how confidence the model is in the prediction. XAI may help develop user trust in the model output [6], where accountability of the model decisions are also required.

In this work, we develop an XAI approach for the analysis of telecommunications call activity that enables greater model inspection and parameter configuration, to account for analyst insight or information unaccounted for by the

Authors' addresses: James Barrett, james6.barrett@uwe.ac.uk, University of the West of England, Bristol, United Kingdom; Phil Legg, phil.legg@uwe.ac.uk, University of the West of England, Bristol, United Kingdom; Jim Smith, james.smith@uwe.ac.uk, University of the West of England, Bristol, United Kingdom; Chip Boyle, cboyle@rbbn.com, Ribbon Communications, Massachusetts, USA.

2024. Manuscript submitted to ACM

model (e.g., external factors). Whilst previous works have explored call data records, there is little research on XAI for telecommunications analysis. We devise a data generation scheme that can model call activity accounting for external influences, and we assess the performance of existing time series forecasting models against an XAI-driven approach. We find that our approach is able to assist an analyst in model tuning and reduce forecasting errors.

The key contributions of this paper are:

- We examine the current research on time series forecasting for anomaly detection.
- We introduce a novel interactive method for configuring a time series ensemble to improve analyst decision-making.
- We conduct an experiment to assess the effectiveness of our approach compared against ML forecasting techniques.

2 RELATED LITERATURE

Our review of the literature will concentrate on interactive time series forecasting and explainable AI techniques.

Spatial-temporal analysis is used in the context of a telecoms fraud investigation by Chen and Shih [2], which closely aligns with the goals of this research paper. The paper involves tracking potential fraudulent suspects through Taiwan and attempts to correlate this data to match fraud suspects with real world movement. The usage of data visualisation via mapping of data points provides a clear advantage to the targeted users of this group in law enforcement, making the process of uncovering connections in the data easier. Although this analysis of user movement in the real world is a useful test case, the author makes the assumption that all data necessary to complete their investigation is within access at all times, it also does not consider any given time-sensitivity, in a real world scenario as an example, data accessibility and time management may become important factors to the law enforcement target audience and therefore should be considered in any similar experiment in the future.

Interestingly, the spectrum of telecom fraud cases is the subject of one research publication [12] that explores the data surrounding groups of users across different age ranges and occupations in Hong Kong and China. This paper makes an excellent effort to explore the dataset used in the initial investigation, it accurately classifies the characteristics of each user group and how they are susceptible to various types of fraud enabled by telecoms. Whilst the exploration and effort to understand the data is thorough, the paper dedicates a small portion to a machine learning test in fraud detection. The paper tests Naive Bayes, Logistic Regression, Support Vector Machines, and Decision Trees, with the last model achieving a 97% accuracy on the dataset, however explains that due to the size of the dataset still results in 300,000 misclassifcations. The paper does explain that a neural network or NLP methods may increase the success of classifying and catching these methods. This research assumes that simple classification in machine learning will suffice in addressing the issue of fraudulent telecom use, whilst also assuming speculating that further policy changes could also aide in fighting these kinds of fraud. It may be noted that more extensive machine learning testing could have supported a stronger solution to the problems highlighted by the data, such as exploring the use of different models as discussed, or potentially including a framework or set of suggestions for researchers exploring the same area, as the results from the machine learning test appear to be only a small part of this paper.

We see time series based machine learning explored in the context of predicting statistics for COVID 19 data in the literature [4], implementing forecasting and "nowcasting" strategies throughout. The strengths of time series forecasting techniques in this paper lie in the methods used to optimise the accuracy and performance of the predictions. This paper implements feature selection whereby the criteria for the features used in training are carefully considered, we also see the usage of feature engineering also paying close attention to the temporal structure of the case metrics used in the data. The key steps of this paper also incorporate a suitable data preparation process, hyperparameter algorithm tuning, performance measuring and model comparison. What may support other research into similar domains in this paper is also Table 1, which features a useful comparison and contrast of different forecasting models and their suitable for "nowcasting" or forecasting. As this publication is comprehensive in scope regarding time series forecasting, it may have been a useful attribute to have made the data used openly available or made the specific parameters used for each model readily available to view so that the experiments could have been reproduced in an easier manner or extrapolated to other public health projects at the time. In the context of our research, we can consider sources such as these and begin to formulate the criteria of a suitable forecasting experiment.

Lewis and Marsh address the challenge of XAI trustworthiness [6]. They look at multiple factors that can influence trust, since trying to establish a single measure of trust can be challenging or ambiguous. The paper interesting makes comparisons between trust and seemingly inanimate objects and animals as a metaphor for increasingly levels of trust between a user a such metaphor. The paper makes an to engage several important conversations in which must be asked when learning to trust systems with increasingly important or sensitive data. This paper opens the conversation, and in terms of ethics, to what level AI should be trusted and how reliable the insight it generates is. Whilst the discussion in this paper is useful, links between real world case studies could help bridge gaps made between hypothetical arguments, as it is often important to link these ideas to examples such as system design whereby it may be useful to understand and study how a user learns to trust the analytics of a system.

3 PROPOSED METHOD

Prior works have explored the development of time series forecasting models, using neural networks for model validation while also fine tuning model hyper parameters [8]. However, in this work we propose an active learning system [5] that allows for the inspection and interrogation of multiple time-series models, enabling an analyst to swap, modify, or update parameters of the model between forecast time steps. In particular, this enables the analyst to further capture their own intuition or knowledge about external factors, and to understand whether anomalies are based on the observed, or unobserved, attributes that impact on the overall telecommunications system.

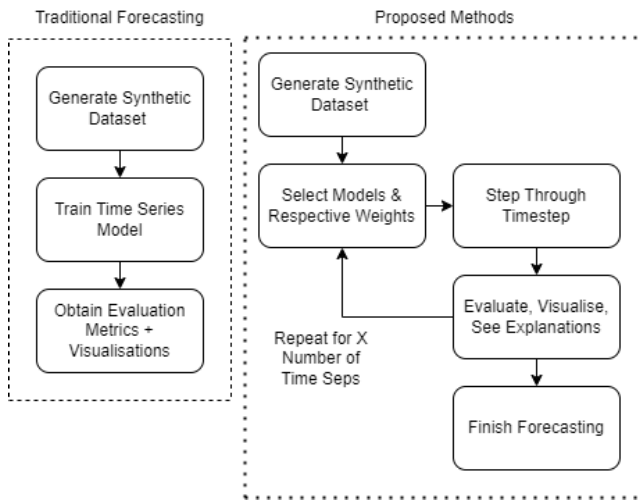

Fig. 1. Diagram of proposed method

Figure 1 demonstrates our proposed method.

Our method of forecasting enables Shapley Explanations (SHAP) among evaluation metrics to determine ensemble effectiveness during forecast windows. We utilise SHAP by creating a waterfall plot of model contribution (negative or positive) between time steps, in which can be used to inform the user, where they may choose to reconfigure model weights, or include/exclude additional models. Implementing SHAP explanations in the near real-time to give the role of analyst greater control of a ML model ensemble enables a method for efficient and dynamic reconfiguration of models between time steps. This not only implements SHAP to ensure efficiency in forecasts, however potentially saves computational resources by disregarding inferior models from being trained.

The experiment will involve a typical time series forecasting scenario with synthetically generated pseudo-random telecoms data that attempts to mimic the conditions of a calls made in these networks, usually recorded in the form of call detail records. The test environment will load a set of pre-programmed scenarios that emulate network service issues in which predictive forecasting is applied. In order to test the difference between our approach and traditional time series forecasting approach, we must build an environment whereby we use the same data and compare the circumstances where we note the difference of our approach. We note below the process for both of our tests.

- Firstly, we generate 2 days of call data with 10,000 users. These profiles are aggregated to make uniformed, grouped data on a minute by minute basis.
- **We then inject 20 external events** that correspond to real world events, that all present in different ways. These are distributed across the 2 days of data.
- The time series data is fed to 10 baseline models, with the last 5 hours being used as test data. Each time step window is 10 minutes.
- Now the data is fed to the interactive method, using the same time steps and test-train period.

Figure 2 shows a 48-hour period of aggregated call counts with a distinct pattern to show the increase and decrease of dropped calls. We also notice variations in the minor fluctuations of this aggregated time series data, although this may appear to relatively random noise, it is influenced by caller profiles with pre-defined characteristics and patterns.

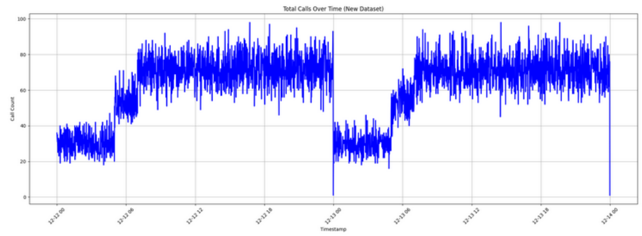

Fig. 2. Visualisation of Time Series Dataset 1.

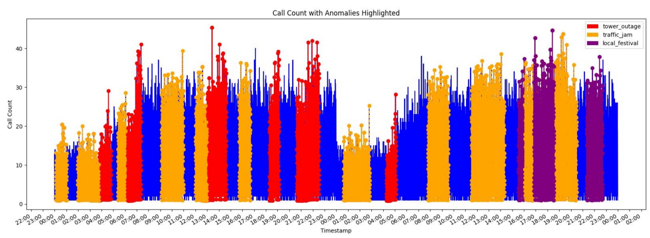

Fig. 3. Visualisation of Time Series Dataset 1 with Real Events Injected.

This data is generated to appear similar day to day (in natural patterns, such as night and day activity), however, callers will not repeat the exact same patterns of the previous day, as the odds of this (depending on size) are statistically near-impossible.

Figure 3 shows the occurrence of event injection mapped against the underlying call activity data. The primary purpose of injecting external events is to expose the time series models to unforeseen events not typical of the baseline time series pattern. These events present using the same parameters but vary and fluctuate subtly. We test our time series data on the last hour of test data, which may or may not contain a whole, or partial event. The logic of this is that in the circumstance an event occurs, the forecaster will be tested to determine how effectively it can adjust to one-off or recurring time series events.

3.1 Data Generation

The time series data generated for this experiment is a synthetically generated version of Call Detail Records (CDRs) as seen in the telecoms industry. These records are responsible for logging information and other metadata [1] on calls and other network transactions that occur over cellular networks. We notice an example of CDR data used for a similar predictive task [10], however CDR data in this case differs from the data we have generated, primarily that the data used for this study is from 2013 [3] and may have be subject to change as telecom data analysis has evolved over the past 10 years, but also that this data is from one operator in one city in a limited time frame. We acknowledge this dataset is a useful reference, however, it does not provide detail of any underlying phenomenon as to why time series events occur in any given manner. Considering this, we therefore decide to take a data generation approach to our work, this populates the series of events that occur within our synthetic environment (being that of a cell tower network), resulting in the traffic of calls occurring over the network, driven by the needs of the users in response to the proposed events. The datasets we generate for the experiments are configured to include 10,000 unique users over the course of 48 hours. The first dataset contains 16,724 rows, the second 17,212, the third 17,147 rows. These rows are aggregated from larger caller profiles, therefore there will be a small row difference between datasets. We then use a separate script to inject 20 anomalous events, these vary from heavy traffic, cellular tower failure, and local festival. These events are set to occur at the following rates: traffic outage (30% likely), heavy traffic (60% likely), local festival (10% likely).

3.2 Forecasting Case Scenarios

We simulate three datasets in this experiment. Each dataset has been generated for the same duration of time (3 years) and number of network users (3000 unique users). We also inject the same amount of events at the same rate to each of these datasets. Therefore, each of the three datasets are similar in nature with similar variations, however vary enough to determine if the forecasting tools perform as intended with this type of data. The generation process involves creating a series of underlying conditions with pseudo-random logic to build baseline patterns, that populate seasonality across different time granularities. Logs are generated in the form of a table that details individual call records, such as one user calling another for a pre-determined amount of time based on the group and characteristics of that user.

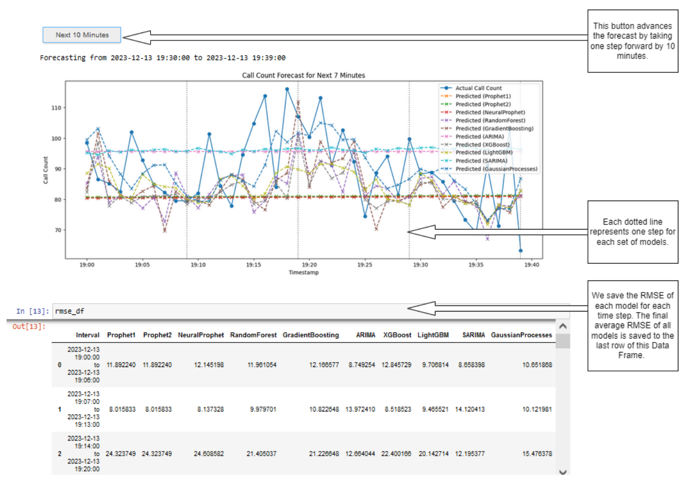

Fig. 4. Example usage of the baseline modelling tool. Compares predictions of the 10 chosen time series forecasting models against the true data value.

4 EXPERIMENTATION

This section covers the methods used to complete the forecasting experiments. Both methods use exactly the same datasets.

4.1 Results From Baseline Model Test

Our results, 1 show an analysis of the state of the art models and methods that are available for the task of time series forecasting and compare each by root mean squared error.

Table 1 shows the RMSE of each baseline model when tested with our data. We notice that the range of RMSE is small, with DS1 having an RMSE range of 9.55 to 10.82, DS2 ranging from 8.03 to 9.04 and DS3 ranging between 9.30 and 10.51. We also note "Prophet1", "Prophet2" and "NeuralProphet" scoring the same RMSE for DS1. It is also observed "Prophet1", "Prophet2" Figure 4 shows the baseline model results within our analysis tool.

Table 1. Baseline Models: RMSE From Datasets 1-3

Model	RMSE (DS1)	RMSE (DS2)	RMSE (DS3)
Prophet1	**9.55**	**8.03**	9.41
Prophet2	**9.55**	**8.03**	9.41
NeuralProphet	**9.55**	8.07	9.53
RandomForest	10.33	8.75	10.14
GradientBoosting	9.62	8.35	**9.30**
ARIMA	9.73	**8.03**	9.33
XGBoost	9.74	8.33	9.45
LightGBM	9.69	8.31	9.48
SARIMA	10.82	8.21	10.27
GaussianProcesses	10.52	9.04	10.51
Best Baseline RMSE	**9.55**	**8.03**	**9.30**

Table 2. Interactive Method: RMSE From All Datasets

Test Number	RMSE (DS1)	RMSE (DS2)	RMSE (DS3)
1	9.67	8.04	**9.27**
2	9.7	**8.00**	**9.27**
3	**9.52**	8.11	9.30
4	9.78	8.03	9.94
5	9.68	8.24	9.46
6	9.51	8.84	**9.27**
Best Interactive RMSE	**9.52**	**8.00**	**9.27**

4.2 Results From Interactive Method

Table 2 demonstrates six tests conducted with a configured ensemble. The first column indicates the position of the test (i.e. the first test is 1) and the remaining columns indicate which dataset obtained the respective RMSE. For instance, to read the results of Dataset 1, you would read down the "RMSE (DS1)" column to see how each test performed. We demonstrate in Table 2 how the interactive method performs against the lowest RMSE of any of the baseline models. We notice, over the 6 configured tests a lower RMSE (0.03) for on DS1, a lower RMSE (0.03) for DS2 and a lower RMSE (0.03) for DS3. Although the improvement is small, no standard state-of-the-art model can outperform our interactive method alone in our six tests of configurations.

Our proposed methods involve a hybrid model approach based on active user reconfiguration to a near real-time data source. Our method utilises the effectiveness of these existing methods, however, our method takes advantage of disregarding ineffective models in an ensemble. We understand the suitability of some models may handle a given data point or prediction differently, however, at other, or perhaps overall, changes in time series data may indicate that the underlying algorithms of these models may render ineffective due to shifts in seasonality or data therefore the dynamic reconfiguration process becomes crucial in reconfiguration.

Figure 5 shows the first 10 time-steps for the interactive ensemble method, achieving a lower RMSE. Our approach incorporates SHAP (Shapley Additive Explanations) [7] along with a time-series plot and configurable weightings of the ensemble methods. By measuring step-by-step metrics, an analyst can adjust the ensemble's weighting based on model performance evaluations.

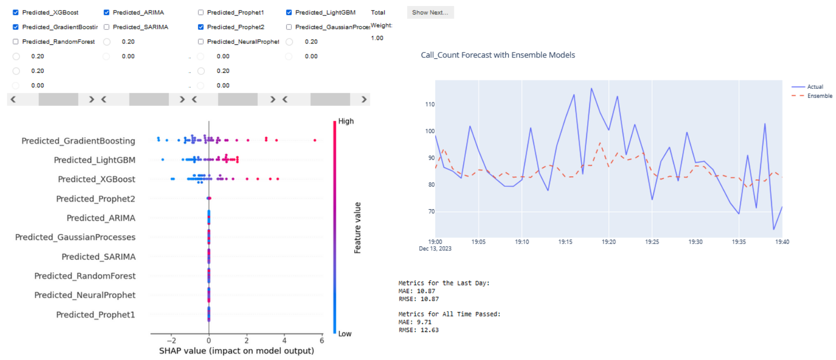

Fig. 5. Depiction of dynamic reconfiguration of models in near real-time forecasting using interactive ensemble techniques.

5 DISCUSSION

Our research demonstrates the effectiveness of an interactive ensemble approach in enhancing time series forecasting accuracy for telecom networks, particularly in near real-time scenarios. This method, integrating human judgement with machine efficiency, surpasses current models but assumes data uniformity. Our future work will explore improving adaptability to varied and dynamic data. The importance of user-friendly interface design is also crucial for users with varying technical backgrounds. Addressing human error in our human-in-the-loop system is essential, as human-machine collaboration, though promising, can introduce biases. This aspect may benefit from a user-based study. Our aim is to optimally utilise both human insight and machine analysis for managing predictions, especially under anomalous or unexpected conditions.

6 CONCLUSION

This paper proposed a new interactive and explainable approach to predictive maintenance for time series data, covering call data records seen in telecoms networks. Data is generated based on real world observations of call detail records, creating profiles based on features seen in this type of data. State-of-the-art models forecast call demand rates and then compared via RMSE to our proposed approach. Comparison between the testing of the baseline model tests demonstrates a lower RMSE than the most accurate state-of-the-art forecasting models for each dataset. Our interactive XAI approach not only outperforms the current best performing time series models, it introduces the capability to perform dynamic model reconfiguration between time steps in near-real time scenarios. The tests performed in these experiments aides demand forecasts in telecoms, whilst also providing a new technical approach for the field of near real-time forecasting where human in-the-loop is considered forefront.

REFERENCES

[1] Zagroz Aziz and Robert Bestak. 2018. Analysis of Call Detail Records of International Voice Traffic in Mobile Networks. In *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*. 475–480. https://doi.org/10.1109/ICUFN.2018.8436669

[2] Ren-Zhi Chen and Chih-Hung Shih. 2021. Applying Spatio-temporal Analysis in Telecom Fraud Investigation. *Procedia Computer Science* 192 (2021), 3109–3113. https://doi.org/10.1016/j.procs.2021.09.083 Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 25th International Conference KES2021.

[3] Telecom Italia. 2015. Telecommunications - SMS, Call, Internet - MI. https://doi.org/10.7910/DVN/EGZHFV

[4] Mallory Lai, Yongtao Cao, Shaun S. Wulff, Timothy J. Robinson, Alexys McGuire, and Bledar Bisha. 2023. A time series based machine learning strategy for wastewater-based forecasting and nowcasting of COVID-19 dynamics. *Science of The Total Environment* 897 (2023), 165105. https://doi.org/10.1016/j.scitotenv.2023.165105

[5] Phil Legg, Jim Smith, and Alexander Downing. 2019. Visual analytics for collaborative human-machine confidence in human-centric active learning tasks. *Human-centric computing and information sciences* 9 (2019), 1–25.

[6] Peter R. Lewis and Stephen Marsh. 2022. What is it like to trust a rock? A functionalist perspective on trust and trustworthiness in artificial intelligence. *Cognitive Systems Research* 72 (2022), 33–49. https://doi.org/10.1016/j.cogsys.2021.11.001

[7] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777.

[8] Md. Mijanur Rahman, Mohammad Shakeri, Fatema Khatun, Sieh Kiong Tiong, Ammar Ahmed Alkahtani, Nurul Asma Samsudin, Nowshad Amin, Jagadeesh Pasupuleti, and Mohammad Kamrul Hasan. 2022. A comprehensive study and performance analysis of deep neural network-based approaches in wind time-series forecasting - Journal of Reliable Intelligent Environments. https://link.springer.com/article/10.1007/s40860-021-00166-x#citeas

[9] Rohit Saluja, Avleen Malhi, Samanta Knapic, Kary Främling, and Cicek Cavdar. 2021. Towards a Rigorous Evaluation of Explainability for Multivariate Time Series. *CoRR* abs/2104.04075 (2021). arXiv:2104.04075 https://arxiv.org/abs/2104.04075

[10] Senem Tanberk and Onur Demir. 2022. Multivariate Modeling and Analysis for Cellular Traffic Prediction Using Call Detail Records. In *2022 7th International Conference on Computer Science and Engineering (UBMK)*. 311–316. https://doi.org/10.1109/UBMK55850.2022.9919559

[11] Xianli Wu, Huchang Liao, and Ming Tang. 2023. Decision making towards large-scale alternatives from multiple online platforms by a multivariate time-series-based method. *Expert Systems with Applications* 212 (2023), 118838. https://doi.org/10.1016/j.eswa.2022.118838

[12] Chunjin Zhu, Chenlu Zhang, Renke Wang, Jingwen Tian, Ruoxuan Hu, Jingtong Zhao, Yaxin Ke, and Ning Liu. 2023. Building of safer urban hubs: Insights from a comparative study on cyber telecom scams and early warning design. *Urban Governance* (2023). https://doi.org/10.1016/j.ugj.2023.05.004

Received 5th January 2024