Forecasting Urban Tourism in Napoli: A Comparative Study of Machine and Deep Learning Models

Ilaria Mangiacapra, Alfredo Nascita, Antonio Montieri, Antonio Pescapè
University of Napoli Federico II (Italy)
{ilaria.mangiacapra, alfredo.nascita, antonio.montieri, pescape}@unina.it

***Abstract*—Accurate and timely forecasting of tourist flows is essential for supporting informed decision-making by local authorities and tourism stakeholders, especially in the absence of up-to-date official data and in the case of exceptional events. This study investigates the application of advanced data-driven models to forecast monthly tourist arrivals at accommodation facilities in the city of Napoli, with a particular focus on the disruptions caused by the COVID-19 pandemic. We perform a systematic comparison of multiple forecasting methods—including traditional machine learning regressors, ensemble models, advanced time series techniques, and deep learning architectures—implemented through a sliding window approach. Experiment evaluation is conducted on both the original time series, which includes pandemic-related distortions, and a reconstructed version in which COVID-era values are replaced using algorithmic interpolation. The results show that reconstructing the pandemic-affected data significantly improves predictive performance, yielding a reduction in forecasting error of over 80% and an increase in the coefficient of determination by more than 0.50. These findings highlight the importance of targeted preprocessing strategies and robust prediction methods for enhancing the accuracy and reliability of tourism demand forecasts.**

***Index Terms*—Tourism flow prediction, Machine Learning, Deep Learning, time-series analysis, COVID-19.**

I. Introduction

Tourism is one of the most dynamic and economically significant sectors worldwide, with substantial impacts on both local development and national GDP. Italy, in particular, stands out for its vast cultural, artistic, natural, and gastronomic heritage, making it one of the most attractive tourist destinations globally [1]. Within this national context, some cities play a pivotal role in attracting tourists; among these, Napoli has emerged as a key destination, having recorded a significant increase in visitor numbers in recent years and strengthening its relevance both nationally and internationally [2].

Hence, the analysis and prediction of tourist flows represent a strategic tool for the planning and management of territorial resources, particularly in areas with a strong tourism vocation. Reliable forecasts are an essential asset for supporting evidence-based decision-making and long-term planning with positive impacts on local economic sustainability [3, 4]. Accurate forecasts support resource allocation, event planning, and infrastructure development, aligning service supply (e.g., accommodation, transport, utilities) with expected demand. However, obtaining reliable forecasts is a complex process due to the integration of information from heterogeneous sources such as accommodation facilities, public transport systems, and local tourism boards.

In recent years, the use of time series forecasting has gained popularity as an effective approach to analyzing the evolution of tourism demand. Indeed, this methodology supports the identification of long-term trends, seasonal components, and anomalous patterns. While numerous predictive techniques have been proposed—ranging from traditional statistical models to advanced *Machine Learning (ML)* and *Deep Learning (DL)* techniques—most of the existing literature focuses on national or regional scales, with limited attention given to *city-level analyses*. At the same time, the *timely availability of tourism data* is essential to inform political, economic, and logistical decisions at all institutional levels. Nevertheless, one of the major obstacles to achieving these goals is the delayed release of official tourism statistics, which are typically published several months or even years after the reference period. This latency reduces institutional responsiveness and limits the ability to take proactive action.

This study aims to address these issues by proposing a forecasting methodology capable of predicting tourist flow data based on historical information. The analysis targets the city of Napoli, using monthly historical data on guest arrivals in accommodation facilities between 2008 and 2023 (those available at the time of writing), sourced from the *Italian National Statistics Institute (ISTAT)*.

A key challenge addressed in this study relates to the COVID-19 pandemic period (2020-2021), during which travel restrictions caused unprecedented disruptions in tourism flows. These disruptions introduced structural breaks and irregularities that complicate both model training and the reliability of forecasts. Additionally, the inherent seasonality of tourist arrivals, characterized by regular peaks and troughs, further

This work was partially conducted within the course “Tecnologie ed Applicazioni per la Trasformazione Digitale” at the University of Napoli Federico II. Particularly, the authors would like to thank Alessia Ferro, Chiara Di Ronza, and Daniela Giacobbe for their collaboration with the preliminary experiments of the study. This work has been partially supported by the Spoke 9 “Digital Society & Smart Cities” of ICSC - Centro Nazionale di Ricerca in High Performance-Computing, Big Data and Quantum Computing, funded by the European Union - NextGenerationEU (PNRR-HPC).

complicates modeling in the presence of shocks or anomalies.

To face these challenges, we develop and evaluate predictive models under different conditions to identify the most effective forecasting strategies.

In detail, the core contributions of this work are as follows:

- We exploit a broad range of ML and DL forecasting techniques, including *Linear Regression*, *Decision Tree*, *K-Nearest Neighbors*, *Random Forest*, *XGBoost*, *Prophet* [5], and *Long Short-Term Memory*, to assess their performance on historical data and forecast tourist flows in Napoli for 2024.
- We take into account the impact of the COVID-19 pandemic on tourism flows and propose a mitigation of COVID-19-related distortions by implementing and comparing multiple data-processing strategies, which include retaining raw data with pandemic years, and reconstructing "normal" trends using algorithmic interpolation [6]. We emphasize that these data-processing strategies are generalizable to any context in which anomalous distortions can be identified in the underlying trends, regardless of the nature of the disruptive event, thus extending beyond the specific case of COVID-19.

The remainder of this manuscript is structured as follows. Section II discusses the main contributions in the literature related to time series-based tourism demand forecasting and positions the present work. Section III details the adopted methodology, including data pre-processing, windowing, and forecasting models. Section IV describes the experimental setup and evaluation metrics. Section V presents the experimental results and related outcomes. Finally, Section VI concludes the paper and outlines directions for future work.

II. Related Work

Forecasting tourism demand has been extensively investigated in the literature, encompassing a wide range of statistical and ML methods. Early approaches predominantly employed econometric and multivariate regression models, leveraging macroeconomic indicators such as income levels, exchange rates, and relative prices to explain fluctuations in tourist flows [7]. Subsequently, autoregressive time series models, such as ARIMA and SARIMA, gained popularity—especially for short-term forecasting—due to their ability to capture seasonal and trend components in historical data [8, 9].

The increasing availability of data and computing power has encouraged the adoption of ML techniques, including Support Vector Regression [10], Random Forest [11], and Deep Neural Networks [12]. These approaches have demonstrated greater flexibility in modeling non-linear relationships and detecting complex temporal dependencies compared to traditional statistical models. In particular, Recurrent Neural Networks (e.g., Long Short-Term Memory and Gated Recurrent Unit) have been successfully applied to tourism time series due to their ability to learn long-term dependencies and adapt to dynamic changes [12]. More complex hybrid or ensemble approaches capable of handling non-linearity and regime shifts have also been explored. For example, in [13], a deep ensemble model combining autoencoders with extreme learning machines proved effective in capturing the complex seasonal and shock-driven dynamics of international tourist arrivals in Beijing.

More recently, researchers have investigated the integration of digital and behavioral data sources, including web search trends, social media activity, and online reviews. These features are often used as proxies for latent tourism demand and can significantly enhance forecasting accuracy when incorporated into hybrid models [11, 14, 15]. However, due to the high collinearity among these features, dimensionality reduction techniques (e.g., principal component analysis and dynamic factor models) are commonly required to improve model performance [16]. Other works [17, 18] have exploited online review text data from platforms like TripAdvisor within mixed data sampling models, combining low-frequency time series with high-frequency user-generated content, and have reported promising results.

The COVID-19 pandemic introduced significant challenges for tourism forecasting by causing structural breaks and adding unpredictable noise to historical time series. Recent studies, such as [19], focused on the Airbnb market in Madrid, observed substantial drops in ML models performance during the pandemic, particularly when COVID-related variables were not explicitly incorporated. To address these limitations, a more recent work [20] has proposed integrating heterogeneous data sources with varying frequencies through mixed data sampling techniques for forecasting tourist arrivals from mainland China to Hong Kong.

Despite the growing body of research, the vast majority of studies are conducted at the national or regional level, with limited attention to the urban scale. Some exceptions include forecasts in metropolitan areas such as Vienna, Puerto Rico, or Hong Kong, which often leverage alternative data sources like Google Trends or social media platforms [20, 21, 22]. However, few studies adopt a comprehensive comparative approach across multiple model families within a city-level context.

A. Positioning of Our Work

In light of related research, our study aims to address two key gaps in the literature. First, we focus on a *fine-grained urban setting*—namely the city of Napoli in Italy—an area that has received limited attention in tourism forecasting research despite its growing importance as a travel destination [2]. Second, we perform a *systematic comparison* across a diverse set of data-driven models from different methodological families, including: (i) traditional ML regressors (Linear Regression, Decision Tree, and K-Nearest Neighbors), (ii) ML ensemble methods (Random Forest, XGBoost), (iii) advanced time series models (Prophet), (iv) and deep recurrent neural networks (Long Short-Term Memory).

Unlike previous works, this comparison is carried out under different data conditions, including both the *original time series* and *reconstructed versions* to account for the

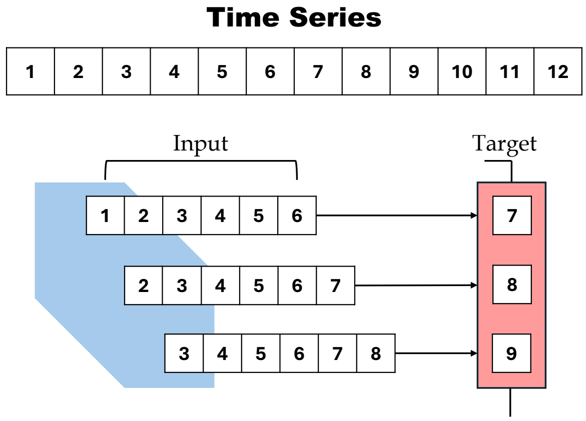

Fig. 1. Example of input-output construction for the prediction task starting from an initial time series. The example shows a sliding window with a fixed size $W = 6$ (*input*); $n+1$ is the index of the value to be predicted (*target*).

pandemic impact and handle COVID-19-induced disruptions. By adopting this *multifaceted approach*, our contribution is both methodological and empirical, aimed at enhancing the robustness and reliability of tourism forecasting in urban planning contexts.

III. Forecasting Methodology

The forecasting methodology adopted in this study is based on the application of prediction techniques to a univariate time series representing monthly tourist arrivals in the city of Napoli. The problem is formulated as a single-step prediction task: the objective is to estimate the number of arrivals for the next month based on the values observed over the previous W months. In Section III-A, we formalize the prediction task and describe the windowing strategy adopted to construct the input sequences for the prediction models. Then, in Section III-B, we provide an overview of the prediction models leveraged in our analysis.

A. Prediction Task and Input-Output Construction

The goal of our prediction task is to forecast the number of tourist arrivals for the following month $n+1$, denoted as $\hat{x}^{n+1}$, based on past observations up to time n. To this end, we define a predictor function $\mathcal{M}(\cdot)$, which takes as input a sequence of W past observations and returns the predicted value for the next time step (viz. month):

$$\hat{x}^{n+1} = \mathcal{M}\left(x^{n-W+1}, \ldots, x^{n-1}, x^{n}\right).$$

To construct the input-output data for the forecasting models, we apply a *sliding window* approach with unit stride, as illustrated in Fig. 1. In this procedure, a fixed-size window of length W is moved across the time series one time step at a time, generating input-output pairs from partially overlapping sequences. This ensures that all training instances contain the same number of time steps W, allowing the models to learn sequential dependencies without processing the entire series at once. With this setup, the earliest prediction can be made at time step $n = W + 1$. As n increases, the window slides forward, always incorporating the W most recent observations to generate new predictions.

Formally, given a univariate time series of length T, the sliding window mechanism produces $T - W$ training samples. Each sample consists of a sequence of W consecutive input values and the corresponding target output. This preprocessing step is consistently applied across all predictive models used in this study.[1]

To generate model forecasts, two alternative approaches are employed. The first constructs each input window W using only real data, ensuring that every prediction is based exclusively on actual historical values. This setup is used primarily to evaluate model performance. The second adopts an *autoregressive forecasting approach*, where predictions are generated iteratively. Starting from the final training window, each newly predicted value is fed back into the input sequence to update the window for the next step. This process continues over the entire forecast horizon, simulating a scenario in which future observations are unavailable and the model must rely on its prior outputs (viz. predicted values). Specifically, the latter approach is used both to reconstruct "normal" values for COVID-19 years and to produce projections for years without available data. Although it may lead to the accumulation of errors over time, it provides valuable insight into the robustness of the models under more realistic deployment conditions.

B. Prediction Models

In this section, we describe the prediction models adopted in our analysis. Specifically, we evaluate a set of *seven data-driven forecasting techniques*, selected from four distinct methodological families: (i) classic ML regressors (*Linear Regression*, *Decision Tree*, and *K-Nearest Neighbors*), (ii) ensemble-based ML models (*Random Forest* and *XGBoost*), (iii) an advanced time series forecasting model (*Prophet*), (iv) and a DL-based recurrent neural network (*Long Short--Term Memory*).

Linear Regression (LR) is a basic statistical method that models the relationship between one or more input variables and a target variable under the assumption of linearity. The model estimates the coefficients of the linear function by minimizing the residual sum of squares between the observed and the predicted values. Owing to its simplicity, interpretability, and computational efficiency, LR is widely adopted as a baseline in predictive modeling tasks. However, its accuracy can significantly deteriorate in the presence of non-linear relationships or when key assumptions—such as linearity, independence of errors, and homoscedasticity—are violated.

Decision Tree (DT) is a non-parametric supervised learning model that recursively partitions the input space by applying hierarchical decision rules based on optimal threshold splits.

[1] Note that this windowing procedure is not required for Prophet [5], which is an additive time series forecasting model that can directly forecast future values from raw time series data without requiring manual windowing.

The structure produced is a tree, where each internal node corresponds to a decision based on a feature value, and each leaf node yields a continuous-valued prediction. For regression tasks, the model selects splits that minimize the variance of the target variable within each partition. DTs are highly interpretable and capable of capturing complex non-linear relationships and interactions among input variables. However, they are prone to overfitting, particularly when grown to large depths without appropriate pruning or regularization techniques.

K-Nearest Neighbors (KNN) is a non-parametric, instance-based learning algorithm that predicts the target value for a new input by averaging the target values of its K nearest neighbors in the feature space. The model does not involve an explicit training phase; instead, it stores the entire training dataset and performs all computations at inference time. Although KNN is simple and effective in capturing local patterns in the data, its predictive performance is highly sensitive to the choice of K, which must be carefully tuned to balance the common bias-variance trade-off.

Random Forest (RF) is an ensemble learning method that addresses the overfitting tendencies of individual DTs by aggregating the predictions of multiple trees, each trained on a different random subset of the data. Specifically, each tree is built using a bootstrap sample of the training set and performs random feature selection at each split, which increases diversity among the trees in the ensemble. The final prediction is obtained by averaging the outputs of all individual trees, commonly resulting in improved accuracy, robustness, and generalization.

XGBoost (short for eXtreme Gradient Boosting) is an ensemble learning algorithm based on the gradient boosting framework, designed for high-performance prediction of continuous numerical targets. It constructs an ensemble of DTs, where each tree is trained sequentially to correct the errors made by the previous ones, effectively minimizing a specified loss function. Unlike traditional bagging methods, gradient boosting optimizes residual errors through gradient descent, allowing the model to capture complex patterns. XGBoost introduces additional improvements, such as regularization, parallelized training, and native handling of missing values, which enhance both accuracy and computational efficiency. It is known for its robustness to overfitting and is widely used in regression tasks involving structured data.

Prophet is a time series forecasting model developed by Meta, designed to handle data exhibiting multiple seasonalities and trend changes [5]. It is based on an additive decomposition model, where non-linear trends are combined with yearly, weekly, and daily seasonal components, as well as user-specified holiday effects. Prophet is particularly effective for time series with strong seasonal patterns and multiple years of historical data. It is robust to missing values, trend shifts, and outliers, and requires minimal manual tuning, making it suitable for applications where domain knowledge may be limited.

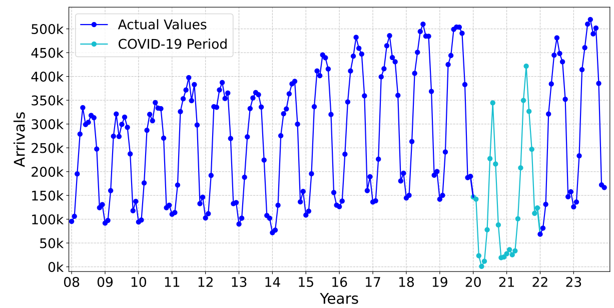

Fig. 2. Monthly tourist arrivals in Napoli (2008–2023), with the COVID-19 period highlighted in light blue.

Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network architecture designed to effectively capture long-term dependencies in sequential data. Unlike traditional recurrent neural networks, LSTM units incorporate internal memory cells along with gating mechanisms (i.e. input, forget, and output gates) that regulate the flow of information. These gates enable the network to selectively retain or discard information over time, allowing it to learn and preserve complex temporal patterns. This capability makes LSTM particularly well-suited for modeling nonlinear, dynamic behaviors often present in time series data.

IV. Experimental Setup

The present section outlines the experimental setup. Section IV-A describes the dataset, while Section IV-B details the evaluation metrics used to assess the performance of the forecasting methodology applied to monthly tourist arrivals.

We also underline that to ensure a fair and unbiased comparison, all prediction models are evaluated using their default hyperparameters as defined by their respective library implementations.

A. Dataset Description

The dataset used in this study consists of a univariate monthly time series representing the number of tourist arrivals at accommodation facilities in the city of Napoli. Tourist arrivals are aggregated from both national and international origins to provide a comprehensive measure of inbound tourism flow. The data, provided by the *Italian National Institute of Statistics (ISTAT)*, cover the period from January 2008 to December 2023, for a total of 192 samples.

Figure 2 depicts the time series, which exhibits a clear seasonal pattern, with recurring peaks in the summer months (July and August), corresponding to the high tourism season, and depressions in the winter months (January and February), typical of the low season. In addition to seasonality, there is a long-term increasing trend interrupted by a sharp and anomalous decline during the COVID-19 pandemic period (2020–2021). Indeed, the global restrictions severely curtailed travel demand and mobility during this period, followed by a gradual recovery in subsequent years. Based on the 192 monthly observations and using a sliding window of length W, the dataset is converted into $192-W$ input-output samples for model training (see Sec. III-A for details).

B. Evaluation Metrics

To assess the performance of the prediction models, we adopt two metrics: the *Root Mean Squared Error* ($RMSE$), defined in Eq. 1, and the *Coefficient of Determination* (R^2), defined in Eq. 2.

Let x_i denote the actual observed values of the target variable, $\hat{x}_i$ the predicted values, and $\bar{x}$ the mean of the observed values. The difference between x_i and $\hat{x}_i$ indicates the prediction error for each observation, serving as the foundation for these performance metrics.

The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \tag{1}$$

It is particularly useful for measuring the discrepancy between predictions and actual observations. Specifically, the RMSE quantifies the standard deviation of the prediction errors; lower values reflect better model performance and a closer match between predicted and actual values.

The R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

The R^2 indicates the proportion of variance in the target variable that is explained by the model. Values approaching 1 denote high explanatory power, while values near 0 indicate a poor fit. Notably, negative R^2 values can occur when the model performs worse than a simple mean-based prediction, indicating that the model fails to capture the underlying structure of the data.

V. Experimental Results

This section presents our experimental evaluation. In detail, Section V-A compares the forecasting results obtained with the original time series, while Section V-B investigates the performance of the prediction models when applied to the data reconstructed to account for the disruptions associated with the COVID-19 period.

A. Forecasting Performance Comparison with Original Data

This section evaluates the prediction models trained on historical data up to the end of 2022 and tested on observations from 2023. To identify the most effective configuration, different sliding window sizes W, equal to 3, 6, and 12 months, are tested for structuring the training set, to determine which setup yields the highest predictive accuracy.

The comparative analysis revealed that the 12-month window obtains the highest performance, indicating that a longer historical horizon contributes to the models' stability and reliability in capturing tourist behavior. Consequently, all subsequent analyses adopt a 12-month window. With this configuration, a total of 180 training instances are generated.

TABLE I
Comparison of different prediction models based on RMSE and R^2 (original dataset).

Regressor	***RMSE***	***R^2***
LR	49092	0.90
DT	41217	0.93
KNN	**32798**	**0.96**
RF	**29950**	**0.96**
XGBoost	48132	0.90
Prophet	114483	0.46
LSTM	40127	0.93

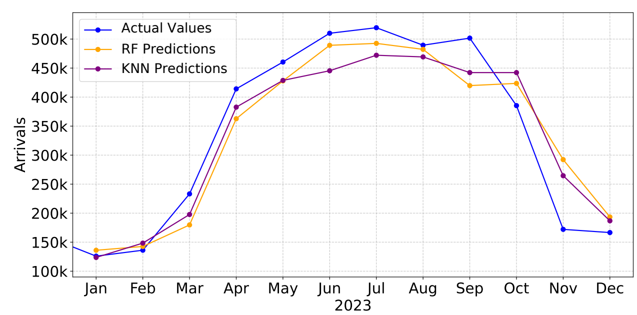

Fig. 3. Graphical comparison of actual monthly arrivals and predictions generated by RF and KNN for 2023 with the original time series.

We recall that Prophet, due to its additive structure and built-in seasonality, generates forecasts directly from the full raw time series, without requiring the sliding window technique for input-output construction.

The forecasting performance varies across models, with some providing predictions that closely match the actual data. The corresponding evaluation metrics are presented in Table I. In detail, KNN and RF achieve the best performance in terms of both RMSE and R^2 values. Specifically, RF exhibits the lowest RMSE of 29 950 and the highest R^2 of 0.96, indicating that this ensemble method is highly effective in capturing complex patterns in the data. Similarly, KNN delivers robust results, with an RMSE of 32 798 and an R^2 of 0.96, showing only a slightly higher RMSE than RF. DT and LSTM also provide competitive performance, achieving an R^2 of 0.93 and RMSE values of 41 217 and 40 127, respectively. These outcomes reflect the models' capacity to capture non-linear relationships, although they are slightly less accurate than the top-performing models. Conversely, Prophet exhibits the weakest performance, with a high RMSE of 114 483 and an R^2 of 0.46, suggesting that it struggles to adapt to the characteristics of the dataset and fails to model the underlying trends effectively—particularly in the presence of anomalous disruptions such as those occurring during the COVID-19 years.

Figure 3 provides a zoomed-in view of the year 2023,

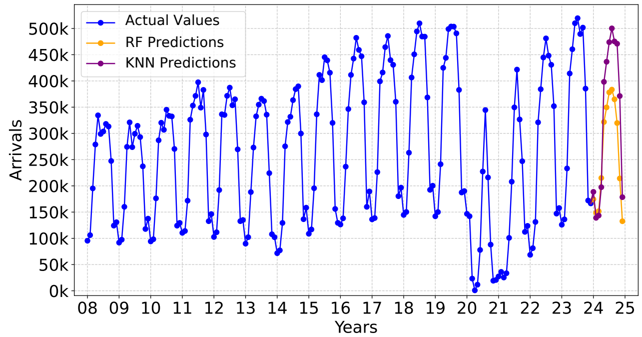

Fig. 4. Forecasts of arrivals for the year 2024 using RF and KNN as prediction models.

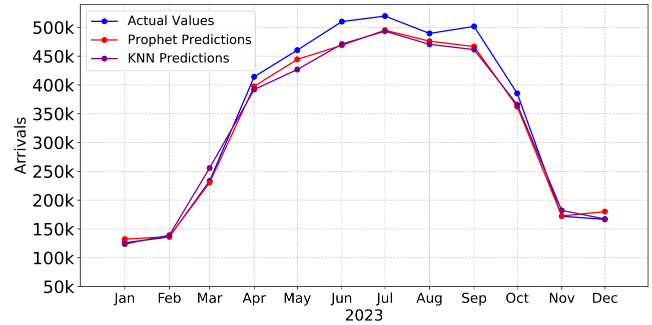

Fig. 5. Graphical comparison of actual monthly arrivals and predictions generated by KNN and Prophet for 2023 with reconstructed time series.

TABLE II
COMPARISON OF DIFFERENT PREDICTION MODELS BASED ON RMSE AND R^2 (RECONSTRUCTED DATA).

Regressor	***RMSE***	***R^2***
LR	43835	0.92
DT	34022	0.95
KNN	**23860**	**0.98**
RF	38759	0.94
XGBoost	32366	0.96
Prophet	**20415**	**0.98**
LSTM	72270	0.78

showing that the best-performing models (i.e. KNN and RF) closely follow the actual trend of tourist arrivals, further confirming the consistency of the quantitative results.

For this reason, these models are selected to generate monthly forecasts of tourist flows for the year 2024. Since actual data for 2024 were not available at the time of writing, the models are trained on historical data up to the end of 2023 and applied using an autoregressive forecasting approach with a sliding 12-month window (see Sec. III-A). In this case, the evaluation is performed visually by comparing the predicted trends with historical patterns to assess their plausibility. The forecasts generated by the models are presented in Fig. 4. We observe that the KNN predictions closely follow the trend observed in recent years (i.e. 2022 and 2023), while the RF model provides forecasts that appear to incorporate longer-term patterns, being more similar to earlier segments of the historical time series.

B. Robust Forecasting with Reconstructed COVID-19 Data

Tourism flows recorded during the years 2020–2021 were profoundly disrupted due to the COVID-19 pandemic, introducing structural discontinuities and significant anomalies in the data. These disruptions compromised the ability of the affected time series to accurately reflect typical historical patterns and seasonal dynamics. This scenario motivated the implementation of a synthetic reconstruction methodology, aimed at evaluating whether the replacement of anomalous data with estimates produced by ML/DL models could improve their predictive performance. To this aim, the models have been trained only using historical data up to December 2019 to prevent bias introduced by the pandemic. It is important to note that conventional metrics for assessing predictive accuracy, such as the RMSE or the R^2, have not been applied in this context, as the aim of the analysis has not been to optimize predictive performance on observed data, but rather to obtain a plausible estimate of the evolution of the analyzed time series in a hypothetical no-pandemic scenario. Each reconstructed time series, one for each model, is used to re-train the corresponding forecasting model (i.e. the same used for the reconstruction), now with data extended up to December 2022, and tested on actual data from 2023 to evaluate the impact of this "correction" strategy on predictive performance.

While evaluation metrics are not computed in the reconstruction phase (2020–2021), they are subsequently applied to evaluate model performance on the 2023 test set, for which actual observations are available. As shown in Tab. II, the best results are achieved by KNN and Prophet, both attaining the highest R^2 of 0.98, and the lowest RMSE values of $23\,860$ and $20\,415$, respectively. DT and RF also obtain competitive results, with R^2 values above 0.94, though slightly less accurate than the top performers. LR showed moderate effectiveness, while LSTM recorded the weakest performance, with both the highest RMSE ($72\,270$) and the lowest R^2 (0.78).

The comparison with results obtained on the original time series (in Tab. I) highlights an overall improvement in performance (with only LSTM showing an evident performance drop) and confirms the benefits of the data reconstruction strategy. Notably, Prophet shows a remarkable enhancement, shifting from the weakest to the best-performing models, with an RMSE reduction of over 80% and an increase in R^2 of more than 0.50. These findings suggest that the reconstructed time series better aligns with the underlying structure assumed by the model. Similarly, KNN and XGBoost also show significant gains, reducing their error by $\approx 27\%$ and $\approx 33\%$, respectively. In contrast, both LSTM and RF do not benefit from the data reconstruction, showing instead a decline in performance. For

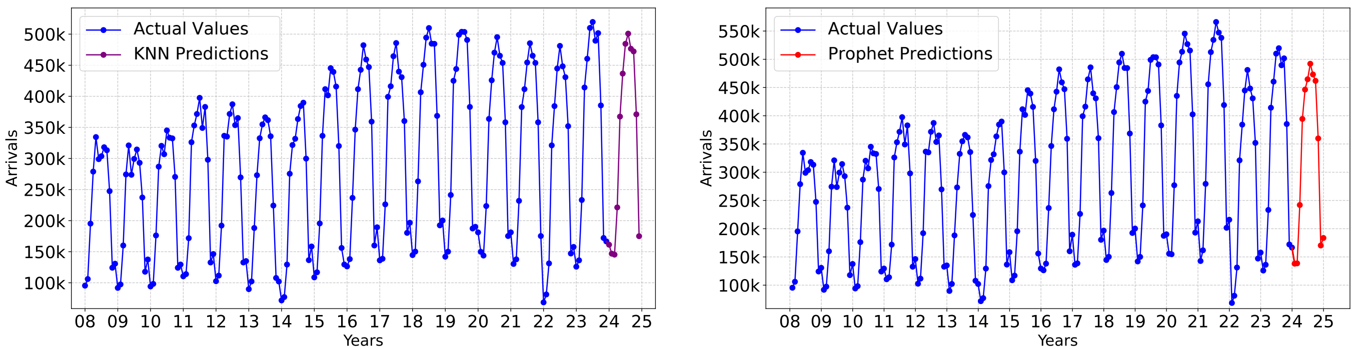

Fig. 6. Forecasts of arrivals for the year 2024 using KNN (left) and Prophet (right).

LSTM, the RMSE increases by $\approx 80\%$ and the R^2 drops by 0.15, while for RF, the RMSE rises by $\approx 29\%$ with a decrease in R^2 of 0.02.

To further support these outcomes, Fig. 5 provides a graphical comparison between the actual monthly arrivals in 2023 and the forecasts produced by the KNN and Prophet models on the reconstructed dataset. Both models closely replicate the seasonal pattern of tourist flows, with only marginal deviations. This evidence further reinforces the effectiveness of the correction strategy in improving the reliability of the forecasts.

Based on these results, KNN and Prophet are selected for further analysis and to generate arrivals forecasts for the year 2024. Specifically, the models are trained with data up to December 2023 and are employed to produce forecasts for the year 2024. Due to the unavailability of official data, the results are visually investigated. Figure 6 illustrates the continuity and seasonality captured by both models in their 2024 forecasts. While slight differences in amplitude are observable, particularly in the summer peaks, both KNN and Prophet maintain reasonable and consistent trends that align with historical seasonal patterns, further supporting the credibility of the pandemic-adjusted strategy.

VI. Conclusion

This study explored the application of advanced data-driven models to forecast tourist flows in the city of Napoli, with a particular focus on the impact of the COVID-19 pandemic. We conducted a systematic comparison of multiple models, including traditional machine learning algorithms, ensemble methods, time series forecasting techniques, and deep learning architectures, trained on historical time-series data. Their performance was assessed on both the original dataset, including COVID-19 distortions, and in a scenario in which "normal" trends were reconstructed using algorithmic interpolation.

The comparative analysis of performance obtained on the original and pandemic-reconstructed time series highlighted that the reconstructed series yields higher forecasting performance. In particular, Prophet and KNN demonstrated improved consistency and reliability in the prediction results. The adopted reconstruction not only mitigated the impact of structural discontinuities but also improved the models' ability to generalize and capture the underlying temporal dynamics. The results indicated that the proposed approach effectively addresses the challenges posed by the inherent variability and seasonality of the tourism data.

Building upon these findings, the future prospects of this work involve both the refinement of the forecasting methodology and the extension of the analysis to broader contexts. A promising direction includes the integration of exogenous variables, such as socio-economic factors, social media activities, or travel trends, which could enhance the models' predictive capabilities. Additionally, exploring probabilistic and ensemble methods may further improve forecast robustness by providing uncertainty intervals for better risk assessment in managing tourist flows. Another promising direction involves the use of Large Language Models for time series forecasting, such as TimeGPT [23]. Thanks to their pre-training on large time-series datasets, these models can enable more flexible and context-aware predictions, potentially complementing or even outperforming statistical and ML-based approaches in capturing complex temporal dependencies. These models are particularly well-suited for forecasting contexts that are highly unpredictable or rapidly changing, such as in the tourism sector, where nonlinear trends and exogenous shocks are frequent. Lastly, applying this forecasting framework to other cities or regions would help assess its applicability through diverse tourism scenarios and provide valuable insights to support policy-making and strategic planning.

References

[1] UNWTO, "International Tourism Highlights," *UNWTO*, 2023. [Online]. Available: https://www.unwto.org

[2] Il Mattino, "Spring 2025: A Record-Breaking Season in Naples," https://www.ilmattino.it/en/spring_2025_a_record_breaking_season_in_naples-8802891.html, 2025, accessed: 2025-06-16.

[3] G. Li, H. Song, and S. F. Witt, "Forecasting Tourism Demand Using Econometric Models," in *Tourism Management Dynamics: Trends, Management and Tools*, 2006, pp. 219–228.

[4] H. Song, R. T. R. Qiu, and J. Park, "A Review of Research on Tourism Demand Forecasting," *Annals of Tourism Research*, vol. 75, pp. 338–362, 2019.

[5] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[6] A. Liu, L. Vici, V. Ramos, S. Giannoni, and A. Blake, "Visitor Arrivals Forecasts Amid COVID-19: A Perspective from the Europe Team," *Annals of Tourism Research*, vol. 88, p. 103182, 2021.

[7] S. F. Witt and C. A. Martin, "Econometric Models for Forecasting International Tourism Demand," *Journal of Travel Research*, vol. 25, no. 3, pp. 306–314, 1987.

[8] F.-L. Chu, "Forecasting Tourism Demand in Asian-Pacific Countries," *Annals of Tourism Research*, vol. 25, no. 3, pp. 597–615, 1998.

[9] C. Lim and M. McAleer, "Time Series Forecasts of International Travel Demand for Australia," *Tourism Management*, vol. 23, no. 4, pp. 389–396, 2002.

[10] K.-Y. Chen and C.-H. Wang, "Support Vector Regression with Genetic Algorithms in Forecasting Tourism Demand," *Tourism Management*, vol. 28, no. 1, pp. 215–226, 2007.

[11] D. Andariesta and M. Wasesa, "Multi-source data integration and machine learning for tourism demand forecasting during covid-19," *Tourism Economics*, 2023.

[12] L. Zhang and Y. Tian, "Forecasting tourism demand in the presence of external shocks using hybrid deep learning models," *Journal of Destination Marketing & Management*, vol. 25, p. 100731, 2022.

[13] Y. Sun, H. Yang, and H. Ma, "B-sake: A hybrid ensemble model for tourism demand forecasting using deep autoencoders and kernel-based methods," *Applied Soft Computing*, vol. 108, p. 107412, 2021.

[14] R. Hewapathirana, "Tourism demand forecasting in sri lanka using machine learning and big data from search engines and social media," *Information Technology & Tourism*, 2023.

[15] B. Pan, D. C. Wu, and H. Song, "Forecasting Hotel Room Demand Using Search Engine Data," *Journal of Hospitality and Tourism Technology*, vol. 3, no. 3, pp. 301–310, 2012.

[16] J. Li, L. Xu, L. Tang, S. Wang, and L. Li, "Big Data in Tourism Research: A Literature Review," *Tourism Management*, vol. 68, pp. 301–323, 2018.

[17] A. F. Colladon, B. Guardabascio, and R. Innarella, "Using social network and semantic analysis to analyze online travel forums and forecast tourism demand," *Decision Support Systems*, vol. 123, p. 113075, 2019.

[18] M. Hu, H. Li, H. Song, X. Li, and R. Law, "Tourism Demand Forecasting Using Tourist-Generated Online Review Data," *Tourism Management*, vol. 90, p. 104490, 2022.

[19] L. Moreno-Izquierdo *et al.*, "Can machine learning improve the forecasting of hotel occupancy rates post-covid-19?" *Tourism Management*, vol. 95, p. 104689, 2023.

[20] J. Wu, M. Li, E. Zhao, S. Sun, and S. Wang, "Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid covid-19? mixed-data sampling approach," *Tourism Management*, vol. 98, p. 104759, 2023.

[21] U. Gunter and I. Önder, "Forecasting City Arrivals with Google Analytics," *Annals of Tourism Research*, vol. 61, pp. 199–212, 2016.

[22] R. Rivera, "A Dynamic Linear Model to Forecast Hotel Registrations in Puerto Rico Using Google Trends Data," *Tourism Management*, vol. 57, pp. 12–20, 2016.

[23] A. Garza, C. Challu, and M. Mergenthaler-Canseco, "Timegpt-1," *arXiv preprint arXiv:2310.03589*, 2023.