Prediction of Hotel Occupancy
LUMEN Data Science

2024.

# Alfatec

FUTURE ENLIGHTENED

# Contents

# 1 About us

## 1.1 Contact information

- Name: Alfatec Group d.o.o.

- Address: Tuškanova 37, 10000, Zagreb

- Contact person: Suzana Srpak

- e-mail: suzana.srpak@alfatec.ai

- Phone: +385 91 572 5657

## 1.2 About Alfatec

For over 30 years, Alfatec Group has been guiding its clients on a journey of change and development, with data at the core of our expertise. As a highly specialized data management company, we're among the leaders of data industry in the Southeast European (SEE) region, supported by our extensive experience with the world's leading ICT technologies and the high levels of satisfaction among clients. Our expertise covers various domains, including IT security, big data analytics, artificial intelligence, data science, business analytics, data storage, databases, and the development of customized software solutions to meet specific user requirements. Our key technology partners include IBM, Thales, SAP, Microsoft, Oracle, Kofax, and Exfo-HS.

Our team of experienced, highly qualified, and dedicated specialists ensures that all our projects are delivered on time, within budget, and to the ultimate satisfaction of our clients. Using cutting-edge AI and ML technology, we innovate and create fast, efficient, and secure products that are easy to use. We have successfully executed over 200 projects and developed advanced solutions. Our experience has shown us that investing in our employees is the foundation of success, and for years, we have been growing our business with a clear vision of mutual growth.

# 2 Challenge description

## 2.1 Intro and motivation

Predicting hotel occupancy is an important issue in the hospitality industry. Hotels and accommodations of all sizes, from small inns to large, luxurious resorts, rely on accurate occupancy forecasts to optimize their operations, revenue, and guest satisfaction. Predicting occupancy levels plays a crucial role in decision-making processes such as staffing, pricing, and resource allocation, ultimately affecting the overall success and profitability of the hotel.

With the advent of data-driven technologies and the increasing importance of online booking platforms, hoteliers have access to a wealth of data sources that can be leveraged for building predictive models. These models enable hotel management to make informed decisions that not only enhance efficiency but also improve the guest experience.

In this task, we will delve into the intricate world of hotel occupancy prediction. We will explore methodologies, data sources, and factors influencing occupancy rates, as well as the benefits of accurate predictions for the hospitality industry. Furthermore, we will examine the challenges and opportunities that hoteliers face in their quest for more precise and insightful occupancy predictions.

## 2.2 Business problem

Welcome to the Hotelier's Challenge! In this thrilling competition, you'll step into the shoes of data analysts for an esteemed hotel chain located on the breathtaking shores of Croatia. Your mission, should you choose to accept it, is to develop a cutting-edge predictive model for hotel occupancy. The hotel chain is counting on your analytical prowess to optimize their resource allocation and revenue management.

Imagine the picturesque coastline, the azure Adriatic Sea, and the enchanting historical sites that make Croatia a sought-after destination. Tourists from neighboring countries flock to this enchanting paradise year-round, seeking the perfect blend of relaxation and adventure. The hotel chain, with its stunning waterfront properties, is at the heart of this travel haven. It's your job to ensure every guest enjoys a memorable stay, while the business thrives.

Your tools? A comprehensive dataset that holds the secrets of the hotel's past. This treasure trove contains historical records of reservations, occupancy rates, and various factors that influence hotel occupancy. But, there's a twist. To enhance the predictive power of your model, you're not just working with internal data.

You'll need to dive into the world beyond the hotel's walls and create a dataset that captures important events from the neighboring countries that frequently send travelers to Croatia.

Here are some guidelines on how to approach this challenge:

1. **Uncover the hidden patterns**: Begin by delving deep into the historical reservation data. Identify patterns, trends, and dependencies in guest behavior. Do they prefer certain days of the week, or do specific months see a surge in reservations? What factors are affecting their decisions? Your data analysis is like detective work, uncovering the secrets of occupancy fluctuations.

2. **The creation of EVENTS table**: You have the power to shape the future dataset. You will design the EVENTS table that captures significant happenings in neighboring countries. These events could include holidays, festivals, major sporting events, or even economic indicators. Your role as a data scientist is to choose the most relevant events that are likely to influence travel decisions. The creation of this dataset is a pivotal step towards supercharging your predictive model.

3. **Time resolutions**: The hotel chain is looking for forecasts at various time resolutions - daily, weekly, monthly, and yearly. The challenge is not just predicting future occupancy but also creating a versatile model that can respond to the needs of different departments within the hotel chain. For example, daily predictions are crucial for the front desk team, while yearly forecasts guide long-term planning and resource allocation.

4. **Data magic**: The magic of data science lies in transforming raw data into actionable insights. You will need to preprocess, clean, and feature engineer the data, making it suitable for the predictive model. The EVENTS table you've created will need to be integrated with the reservation data to unlock its full potential.

5. **Model mastery**: Here's where your true expertise shines. Choose and develop the most suitable machine learning or statistical models to predict hotel occupancy. You'll need to train your model on historical data and validate its performance to ensure it's ready for the real-world challenges.

6. **Real-world Impact**: Keep in mind that your predictions will directly impact the hotel's operations. Accurate forecasts will help the hotel allocate resources efficiently, staff appropriately, and maximize revenue. Think of your work as the engine that drives the hotel's success.

7. **Prediction intervals and visualization**: In the world of data science, predictions alone are not enough. For every forecast you make, it's essential to provide a corresponding prediction interval (or prediction sets). These intervals represent the range within which you expect the actual occupancy to fall, considering the inherent uncertainty in the data and the predictive model. In the quest to make your occupancy forecasts even more robust, you will leverage the power of Conformal Prediction.

As you embark on this data adventure, remember that your journey isn't just about numbers and algorithms; it's about telling a compelling story with data. It's about creating unforgettable experiences for travelers, enabling the hotel to thrive, and ultimately, preserving the magic of Croatia's stunning coastline. So, get ready to dive into the data, bring your creative thinking, and craft a predictive model that'll leave a mark on the world of hospitality. Good luck, data scientist, and may your insights shine as brightly as the Adriatic sun!

## 2.3 Task

Develop a predictive model for hotel occupancy for a Croatian hotel chain. Analyze historical reservation data to uncover patterns and dependencies in guest behavior. Create an EVENTS table to capture relevant events from neighboring countries that influence travel decisions. Provide forecasts at different time resolutions (daily, weekly, monthly, yearly) and integrate the EVENTS table with reservation data. Choose and develop suitable machine learning or statistical models, validate their performance, and ensure the predictions directly impact hotel operations, enabling efficient resource allocation and revenue maximization. Include prediction intervals for added robustness using conformal prediction. The final product should encompass daily, weekly, monthly, and yearly predictions to facilitate planning across various time resolutions.

## 2.4 Dataset description

The primary dataset includes the columns as presented in Table 2, accompanied by example values. The reservation table is a fundamental component of the data in the context of the hotel occupancy forecasting challenge. This table contains structured information about the hotel reservations, and it is a key data source for developing predictive models for estimating hotel occupancy.

The events table shown in Table 1 is envisioned as an additional data source to aid in the estimation, in addition to the existing reservation data. This table needs to be created and filled with data as indicated in the example.

Once you gather and structure this data, you can use it to build predictive models for forecasting hotel occupancy. The quality and relevance of the data you collect will significantly impact the accuracy of your predictions. Don't forget to explore how additional data influences the prediction.

| EVENTS | | | |
|---|---|---|---|
| zemlja_događaja | HR | HR | SLO |
| datum_početka | 08.6.2023. | 11.4.2023. | 22.6.2023. |
| datum_kraja | 08.6.2023. | 17.4.2023. | 22.6.2023. |
| opis_događaja | Tijelovo | Školski praznici Uskrs | Dan bla bla |
| vrijedi_od | 1.1.1990. | | 1.1.1990. |
| vrijedi_do | 31.12.2099. | | 31.12.2099. |
| ponavljanje | NE | NE | DA |
| kategorija_događaja | Tijelovo | Školski praznci | Blabla |

Table 1: Events table containing a record of important events from surrounding countries.

| rezervacija_id | 136929 | 136929 | 134963 | 96205 | 96205 |
|---|---|---|---|---|---|
| gost_id | 0 | 1 | 2 | 3 | 4 |
| zemlja_gosta | HU | HU | HR | HR | HR |
| datum_kreiranja_rezervacije | 2.8.2012 | 2.8.2012 | 1.8.2012 | 27.6.2012 | 27.6.2012 |
| datum_nocenja | 6.8.2012 | 7.8.2012 | 6.8.2012 | 7.8.2012 | 8.8.2012 |
| datum_otkazivanja_rezervacije | NULL | NULL | NULL | 7.7.2012 | 7.7.2012 |
| broj_rezerviranih_soba | 1 | 1 | 1 | 1 | 1 |
| broj_odraslih_gostiju | 2 | 2 | 2 | 2 | 2 |
| broj_djece_gostiju | 2 | 2 | 0 | 0 | 0 |
| cijena_nocenja | 231,900 | 231,900 | 216,630 | 283,790 | 283,790 |
| tip_sobe_id | 31 | 31 | 93 | 80 | 80 |
| kanal_prodaje_id | 5 | 5 | 1 | 1 | 1 |

Table 2: Sample data from the EVENTS table.

# 3 Recommendations for solving the challenge

## 3.1 Technology

- **Python**: Python is an extremely popular language for data analysis and machine learning. You can use libraries such as NumPy, pandas, scikit-learn, and TensorFlow/PyTorch for data processing and model development.

- **Jupyter Notebook**: An excellent tool for interactive data exploration, visualization, and writing Python code.

- **Tableau ili Power BI**: For data visualization and creating interactive charts and reports.

- **Key Terms**: forecasting, conformal prediction, uncertainty estimation for predictions.

## 3.2 Additional literature

Tools and literature that might be helpful: scikit-learn [1], conformal prediction links and examples [2], MAPIE library [3], matplotlib [4], pandas [5], Google Calendar [6], FastAPI [7].

# 4 Solution format

We grant contestants the liberty to select the format of their input and output variables as they see fit. However, we emphasize the importance of thorough documentation. This documentation is crucial to our ability to replicate and understand the results achieved during the competition. It ensures that the methods, processes,

and data used in the competition are transparent and comprehensible to the judges or anyone reviewing the work.

Participants are required to submit their final solution in the form of a well-organized zip archive. This archive should include project documentation and technical documentation in PDF format, the source code, and all the necessary components for deploying the application.

## 4.1  Project documentation in PDF format

Project documentation serves as a detailed record of the entire project. It should encapsulate the project's objectives, methodologies, processes, and any pertinent findings. This document should not only provide insights into the project's goals and how they were achieved but also offer clarity to evaluators and stakeholders regarding the project's overall scope and impact.

## 4.2  Technical documentation in PDF format

The technical documentation, also provided in PDF format, offers an in-depth view of the solution's underlying technical aspects. This may encompass architectural diagrams, data flow charts, algorithms used, and any technical nuances. Clear and precise technical documentation is essential for comprehending the inner workings of the solution, making it easier for reviewers to evaluate the technical soundness and efficiency of the project.

## 4.3  Source code

Participants are expected to include the complete source code of their solution. This entails all programming scripts, source files, and associated assets. The source code is fundamental for understanding the logic and functionality of the solution. It also permits reviewers to asses the quality of the code, adherence to best practices, and its potential for scalability and maintainability.

## 4.4  Code and other necessary elements for deploying the application

In addition to the source code, participants must provide any auxiliary materials required for deploying the application. This encompasses configuration files, libraries, dependencies, or any other elements necessary to make the solution operational.

## 4.5  Presentation

The top teams after the first part of the competition will be invited to the finals, for which they will need to prepare a 15-minute presentation.

## 5  Scoring criteria

| Criteria | Scoring range | Scoring contribution |
|---|---|---|
| Project documentation | 0-10 | 20% |
| Technical quality | 0-10 | 20% |
| Feasibility of the solution | 0-10 | 25% |
| Model Precision | 0-10 | 10% |
| Deployment | 0-10 | 10% |
| Presentation | 0-10 | 15% |

- **Project documentation**: The overall experience of the documentation is being assessed. This includes a general understanding of the problem, the implemented solution, and how it is presented and explained. This encompasses various visuals, diagrams, and examples.

- **Technical quality**: The technical execution (code structure and organization) of the solution is being evaluated, along with the accompanying technical documentation.

- **Feasibility of the solution**: The approach and method of implementing the solution are being assessed, as well as its creativity, the balance of quality and size, and its robustness when adding new features or making minor changes.

- **Model Precision**: Evaluating model precision in occupancy forecasting involves assessing how well the model's predictions match the actual occupancy values by assessing whether the actual occupancy falls within the predicted intervals with the expected confidence level

- **Deployment**: Teams are expected to have an API for testing the solution, but any additional effort invested in a user interface will be rewarded additionally.

- **Presentation**: The evaluation of presentations depends on the visual aspect of the slides, the amount of essential information presented, the clarity of the presentation, and the overall performance.

# 6    Additional benefits

The most qualified candidates are eligible for a paid internship opportunity.

# References

[1] scikit-learn. https://scikit-learn.org/stable/.

[2] Awesome conformal prediction. https://github.com/valeman/awesome-conformal-prediction.

[3] Mapie. https://github.com/scikit-learn-contrib/MAPIE.

[4] matplotlib. https://matplotlib.org/.

[5] pandas. https://pandas.pydata.org/.

[6] Google calendar. https://calendar.google.com/calendar.

[7] Fastapi. https://fastapi.tiangolo.com/.